

Enjeux numériques



Quelles infrastructures numériques du futur ?

N°27 - SEPTEMBRE 2024



*Publiées avec le soutien
de l'Institut Mines-Télécom*

UNE SÉRIE DES

ANNALES DES MINES

FONDÉES EN 1794

ENJEUX NUMÉRIQUES

ISSN 2781-1263 (en ligne)

ISSN 2607-9984 (imprimé)

Série trimestrielle - N°27 - Septembre 2024

Rédaction

Conseil général de l'Économie
Ministère de l'Économie,
des Finances
et de la Souveraineté
industrielle et numérique
120, rue de Bercy
Télédoc 797
75572 Paris Cedex 12
Tél. : 01 53 18 52 68
<http://www.annales.org>

Grégoire Postel-Vinay

Rédacteur en chef

Alexia Kappelmann

Secrétaire générale

Daniel Boula

Secrétaire général adjoint

Magali Gimon

Assistante de rédaction
et Maquettiste

Frédérique Linqué

Webmestre et Maquettiste

Publication

Photos de couverture

Saint Georges terrassant
le dragon, © Ville de Grenoble /
Musée de Grenoble-J.L. Lacroix -
Domaine public

Iconographie

Daniel Boula

Mise en page

Magali Gimon

Impression

Dupliprint Mayenne

Membres du Comité de rédaction

Pierre Bonis

Co-président

Anne-Lise Thouroude

Co-présidente

Edmond Baranes

Godefroy Beauvallet

Côme Berbain

Hélène Brisset

Serge Catoire

Nicolas Chagny

Jean-Pierre Dardayrol

Éric Freyssinet

Frédéric Garcia

Francis Jutand

Arnaud de La Fortelle

Caroline Leboucher

Bertrand Pailhès

Grégoire Postel-Vinay

Maurice Ronai

André Schwob

Laurent Toutain

Benjamin Vignard

La mention au regard de certaines illustrations du sigle « D. R. » correspond à des documents ou photographies pour lesquels nos recherches d'ayants droit ou d'héritiers se sont avérées infructueuses.

Le contenu des articles n'engage que la seule responsabilité de leurs auteurs.

Quelles infrastructures numériques du futur ?

- 05 **Introduction**
Francis JUTAND et Daniel KOFMAN

ÉLÉMENTS TRANSVERSAUX SUR LA CONVERGENCE

- 09 **Des services réseaux aux plateformes du numérique
et aux services multisectoriels**
Francis JUTAND et Daniel KOFMAN

- 15 **Le *cloud* dans la chaîne
de valeur industrielle du numérique**
Quentin ADAM

- 25 **De la manufacture à l'infrastructure :
des conséquences pour la régulation**
Pierre-Jean BENGHOZI

- 31 **Convergence des infrastructures numériques :
un point de vue économique**
Laurent BENZONI

- 40 **De l'eau dans les nuages**
Sylvain BOUVERET, Aurélie BUGEAU
Anne-Cécile ORGERIE et Sophie QUINTON

- 48 **Une perspective numérique européenne**
Emmanuel DOTARO

LES TECHNOLOGIES ET ARCHITECTURES DES INFRASTRUCTURES DU FUTUR

- 58 **La 6G : promesses et défis à l'horizon 2030**
David GESBERT

- 64 **Réseaux de communications
sur constellations de satellites**
François BACCELLI

- 72 **Maîtrise de la complexité dans la 6G -
L'approche par la géométrie stochastique**
François BACCELLI

- 77 **Systèmes cyber-physiques et jumeaux numériques,
déclencheurs et conditions d'une adoption accélérée**
Pascal BROSSET

- 84 **Les apports de Gaia-X**
Anne-Sophie TAILLANDIER et Pierre GRONLIER
- 94 **Les réseaux virtualisés : promesses et enjeux**
Fabrice GUILLEMIN
- 99 **La *softwarization* des réseaux
et son impact technico-économique**
Emmanuel PUIG
- 104 **Nouvelles technologies optiques,
quel impact sur les systèmes
de réalité augmentée de demain ?**
Jean-Louis de BOUGRENET DE LA TOCNAYE
- 113 **Métavers : au-delà des « casques de réalité virtuelle »**
Guillaume MOREAU
- 119 **La cybersécurité**
Hervé DEBAR
- 127 **Les composants pour les infrastructures numériques**
Dimitri KTÉNAS et Dr Denis DUTOIT

LES IMPACTS SECTORIELS

- 134 **Apport de l'infrastructure de partage de données
Agdatahub dans le secteur agri-agro**
Sébastien PICARDAT
- 142 **Les enjeux du numérique
dans la distribution d'électricité**
Claire WAAST-RICHARD, Claudine RABILLARD
et Yves BARLIER
- 149 **Les systèmes de transports intelligents coopératifs
au service des politiques de mobilité urbaine**
Éric MONCEYRON
- 161 **Plaidoyer pour les grandes oubliées :
les infrastructures publiques de partage de données**
Laura LÉTOURNEAU
- 173 **Les systèmes de santé**
Michel BARTH
- 184 **L'évolution du commerce grâce à la numérisation**
Leonardo ÁLVAREZ

ÉLÉMENTS PROSPECTIFS

- 192 Le rôle de l'IA dans l'évolution
des infrastructures numériques du futur**
Francis JUTAND et Daniel KOFMAN
- 199 Du « sens » dans les réseaux**
Laurent CLAVIER et Marios KOUNTOURIS

205 Traductions des résumés

213 Biographies des auteurs

*Ce numéro a été coordonné par
Francis JUTAND et Daniel KOFMAN*

Introduction

Par Francis JUTAND

Conseil général de l'Économie

Et Daniel KOFMAN

Télécom Paris

Les évolutions technologiques rapides et disruptives provoquent des transformations structurelles des infrastructures numériques et ouvrent de multiples opportunités pour l'ensemble des secteurs d'activité, dont l'émergence de nouveaux paradigmes de services et d'applications.

Le développement d'une vision stratégique et la prise de décisions informées requièrent de comprendre à la fois les évolutions des technologies, celles des infrastructures matérielles et logicielles sous-jacentes, ainsi que les nouvelles architectures fonctionnelles et applicatives qu'elles suscitent.

Ces transformations doivent être envisagées de manière holistique avec celles de la sécurité et de la sûreté, de l'impact environnemental et de l'acceptabilité sociale. L'évolution induite sur la géographie des filières doit être analysée, en intégrant les bouleversements géoéconomiques et géopolitiques, notamment dans un objectif de plus en plus pressant de souveraineté.

Ces concepts ont fourni le fil conducteur pour la conception et la rédaction de ce numéro spécial, qui nous l'espérons apportera des éléments pour construire une vision d'ensemble du domaine, de ces enjeux et de ces challenges.

Ce numéro de la série *Enjeux numériques* des *Annales des Mines* est consacré aux infrastructures numériques du futur. Le choix du sujet est motivé, d'une part, par les transformations structurelles en cours du secteur rendues possibles par des évolutions technologiques de plus en plus rapides et disruptives et, d'autre part, par les multiples opportunités que ces évolutions ouvrent pour l'ensemble des secteurs d'activité, dont l'émergence de nouveaux paradigmes de services et d'applications, et la prise en compte de risques majeurs induits et les réponses pour les réduire.

Le développement d'une vision stratégique et la prise de décisions informées requièrent de comprendre à la fois les évolutions des technologies et des infrastructures matérielles et logicielles sous-jacentes, ainsi que les nouvelles architectures fonctionnelles et applicatives qu'elles suscitent. Ce, dans une période où les transformations du numérique doivent être envisagées de manière holistique avec celles de la sécurité et de la sûreté, de l'impact environnemental et de l'acceptabilité sociale. L'évolution induite sur la géographie des filières doit être analysée, en intégrant les bouleversements géoéconomiques et géopolitiques qui impactent fortement les acteurs, notamment dans un objectif de plus en plus pressant et urgent de souveraineté.

La métamorphose numérique a été enclenchée dans les années 1980 avec l'avènement des premiers réseaux numériques à intégration de services et la distribution du calcul rendue possible notamment par l'émergence des serveurs informatiques, des ordinateurs personnels et des réseaux dits locaux. Cela a été suivi par l'introduction de la mobilité et par le changement de paradigme apporté par le *cloud*, engendrant une explosion de nouveaux services, applications et usages. Depuis plus de 10 ans, la convergence entre les réseaux

et le *cloud* a commencé, sur la base de la généralisation des concepts de *softwarisation* et de virtualisation. En même temps, les données se sont érigées comme un vecteur majeur de création de valeur, ce qui se traduit, entre autres, par l'impact croissant et majeur de l'intelligence artificielle, notamment générative.

L'utilisation du numérique : communication, services, applications, intelligence, se traduit en continu dans les entreprises, par la numérisation de fonctions existantes de production, de mesure, de contrôle et de commande distants, et par la création de nouvelles fonctions et de nouvelles organisations. La place émergente mais fortement croissante de la réalité virtuelle, augmentée et mixte, ainsi que celle des jumeaux numériques, attestent de l'ampleur des changements en cours et à venir. Le numérique se situe aujourd'hui au cœur de l'évolution de tous les métiers, de tous les secteurs d'activité et de toutes les tailles d'entreprise.

Les infrastructures de communication, de service de données et d'intelligence se sont développées historiquement en strates avec l'apparition successive de nouveaux acteurs (du *cloud*, mais aussi de la cybersécurité, de l'édition, des applicatifs, etc.) et l'évolution des activités classiques des équipementiers et des opérateurs de Télécom. Les grands acteurs de chaque strate définissent leur offre, avec l'émergence dans les interstices et les niches de nouveaux acteurs, mais aussi explorent des mouvements de conquête sur les chaînes de valeur.

Cette sédimentation en couches se voit aujourd'hui ébranlée par les évolutions du *cloud* qui sort des grands centres de données pour se disséminer sur l'ensemble des infrastructures numériques, citons notamment les concepts de mobile (multi-access) *edge computing* et de *cloudRAN*. Par ailleurs, les architectures normalisées pour la 5G dites "stand alone" permettent l'ouverture du cœur de réseau, permettant à de multiples acteurs d'interagir avec le contrôle des réseaux et à terme d'y déposer leurs propres fonctionnalités.

Pour la première fois dans l'évolution des générations de réseaux mobiles, la 5G a été conçue de manière partiellement synergique avec divers secteurs verticaux, ce qui rend possible cette accélération de la pénétration du numérique au cœur des process des entreprises, notamment industrielles et de réseaux.

La 5G permet de créer de multiples réseaux logiques indépendants grâce à la notion de *slicing*, elle porte également la notion de réseaux 5G privés, permettant aux entreprises de déployer leurs propres réseaux 5G, des réseaux adaptés aux besoins notamment industriels, grâce à la faible latence proposée.

Les réseaux non-terrestres apportent des solutions aux problèmes de couverture, mais auront un impact qui va bien au-delà. Leur interopérabilité avec les réseaux terrestres et les transformations que cela induira chez les opérateurs historiques doivent être évaluées.

Tout ceci sera accéléré avec la 6G et l'avènement de nouvelles technologies, telles que les réseaux sans cellules (*cell-free*), les surfaces intelligentes reconfigurables (RIS) et l'utilisation de parties plus hautes du spectre des fréquences radio.

La collecte, la circulation, le stockage et le traitement, de plus en plus distribués, des données de toutes natures, synchrones ou asynchrones, structurées ou non structurés, demandent de nouveaux services de partage, d'échange ou de fédération, avec des fonctionnalités spécifiques de contrôle de l'accès, de l'utilisation et de traçabilité. Ils ouvrent tout l'éventail de la localisation des calculs ou computations, intégrant des solutions mixtes centralisées et localisées, dynamiquement organisées.

La *softwarisation* croissante des infrastructures, permet de remettre en cause les sédimentations existantes : communication, donnée et calcul, lesquelles peuvent maintenant s'interpénétrer et donner lieu à des systèmes dynamiques et auto-organisés sur des infrastructures plus facilement partageables. Il s'agit donc d'un potentiel *big bang* auquel

nous devons faire face, réseaux de communication à la demande, applications temps réel, calcul et intelligence répartis, mais aussi de partage d'infrastructures, d'orchestrations multisectorielles de nouveaux services, de création de solutions dynamiques, de relations complexes de financement entre l'investissement et l'usage.

Par ailleurs, la complexité des nouveaux systèmes explose et requiert de nouveaux paradigmes pour la planification, le déploiement, l'opération, la gestion et la maintenance de ces systèmes. L'intelligence artificielle va jouer ici aussi un rôle de plus en plus central.

Le potentiel d'évolution est à apprécier, d'une part, du fait de l'environnement intensif en capital pour les infrastructures et ouvert à la concurrence et aux coopérations et, d'autre part, d'un univers d'innovations technologiques et de services loin d'être stabilisé à l'horizon de moyen et long terme. Il faut par ailleurs prendre en compte les aspects géopolitique et géoéconomique de la lutte pour le contrôle et le partage intelligent de la valeur ajoutée, mais aussi des conflits d'investissements entre les grandes compagnies mondiales de services numériques et les opérateurs et équipementiers.

L'horizon temporel pour l'éclairage prospectif est bien sûr inégal tant les composantes évolutives opèrent à des échelles de temps variées. De nouvelles technologies, notamment de réseaux sémantiques, de calcul et réseaux quantiques peuvent venir modeler ces évolutions ou constituer un deuxième ressort d'évolution.

L'idée est donc dans ce numéro de construire une vision holistique présentant chacune des composantes technologiques, systémiques et applicatives, et leurs évolutions, interactions et interpénétrations. Cela est accompagné par une vision du positionnement des acteurs et de leurs évolutions, avec la mise en évidence d'opportunités majeures pour chacun et pour la société en général, et des risques induits, notamment pour la souveraineté nationale. Ainsi, l'état de l'art et les problématiques majeures qui se posent aujourd'hui sont décrits et, quand cela est possible, une projection de moyen/long terme est fournie, même si la complexité des évolutions et les transformations des acteurs en jeu, rend cet exercice périlleux, du fait de la complexification des facteurs et de leurs interactions, ainsi que de la multiplicité des scénarios futurs envisageables.

Nous avons donc fait appel à des acteurs de divers horizons (entreprises, académiques, bureaux d'études et experts, en cœur de métier mais également de divers acteurs de marchés verticaux) afin de couvrir le volet « éléments transversaux », le volet « technologies et architectures », le volet « impact sectoriel » et enfin le volet « technologies de ruptures » qui composent ce numéro spécial.

Le premier chapitre traite donc des grandes tendances ou éclairages transversaux :

- l'intégration numérique et sectorielle des systèmes de communication ;
- le *cloud* dans la chaîne de valeur industrielle du numérique ;
- les éléments de géopolitique et de régulation ;
- les données économiques de l'évolution des infrastructures ;
- des impacts écologiques ;
- la politique européenne.

Le second chapitre traite des technologies qui vont agir en levier et facteurs d'innovation :

- les promesses de la 6G ;
- les communications satellitaires ;
- la complexité et géométrie stochastique ;
- les systèmes cyber-physiques ;
- les apports de Gaia-X ;
- la virtualisation ;
- la *softwarisation* ;
- l'optique et les terminaux de réalité virtuelle et augmentée ;

- le *metavers* ;
- la cybersécurité ;
- les composants.

Le troisième chapitre traite d'impacts sectoriels :

- le secteur agricole à l'ère de la donnée ;
- le numérique et la distribution d'électricité ;
- les systèmes de transport intelligent coopératifs ;
- les infrastructures publiques de données ;
- le système de santé ;
- le commerce électronique.

Le quatrième chapitre traite de deux technologies de ruptures :

- l'intelligence artificielle ;
- les réseaux sémantiques.

Le choix de n'aborder dans ce numéro que le champ des infrastructures est forcément limitatif pour la perception des impacts majeurs des usages et des attendus éthiques du développement du numérique, même si les contributions sur les secteurs applicatifs pallient partiellement cette limitation, mais le sujet en lui-même est déjà suffisamment complexe.

Nous n'avons donc pas prévu une approche géopolitique globale du développement du secteur numérique, ni une analyse historique étendue qui permettrait de comprendre en profondeur les retards industriels européens d'aujourd'hui dans le domaine du *cloud*, des réseaux sociaux, des moteurs de recherche, des places de marché, de l'IA générative et de l'édition logicielle.

Il est donc nécessaire de construire des éléments de souveraineté dans ces domaines, notamment autour de quelques grands acteurs et des réseaux de PME et ETI très actives, et pour cela de garder globalement la main sur le secteur des télécoms, progresser dans le secteur du *cloud* et prendre le train en marche dans le domaine de l'intelligence numérique ou de l'IA et des IA génératives, pour accélérer par la suite grâce à un écosystème industriel comportant un réseau d'entreprises bien positionnées au niveau international sur divers secteurs d'activité.

Par ailleurs, l'environnement numérique et cognitif que tissent autour de l'humain les technologies numériques : services et applications, robotique, automates et drones, interfaces intelligentes, et les problèmes de confiance liés à leurs conceptions et au risque de leur manipulation géopolitique, mercantile ou hostile est seulement partiellement présenté et constitue également un sujet qui doit éclairer le développement des infrastructures et qui méritera des développements plus approfondis.

Des services réseaux aux plateformes du numérique et aux services multisectoriels

Par Francis JUTAND
Conseil général de l'Économie

Et Daniel KOFMAN
Télécom Paris

Nous rentrons dans une nouvelle phase de la transformation numérique. Dans la continuité d'une transformation en silos, par secteur d'activité, nous passons à une phase qui ouvre la voie aux services multisectoriels, transversaux aux divers marchés verticaux. L'énorme potentiel de création de valeur est identifié et ne cessera de se développer vu les possibilités d'innovation qui sont ainsi ouvertes ; les risques pour les acteurs et pour la souveraineté des pays sont multiples.

Cette évolution est rendue possible par une série de nouveaux paradigmes, dont la *softwarisation* et virtualisation des infrastructures, la convergence progressive entre les réseaux et le *cloud*, l'évolution des interfaces, les jumeaux numériques et l'IA, ainsi que l'ouverture des réseaux et systèmes aux acteurs externes. Ce dernier point transforme les infrastructures du numérique dans des plateformes de services, permettant à ces derniers d'être conçus et établis en temps réel, en orchestrant des composants d'acteurs multiples, pour répondre dynamiquement à des besoins spécifiques et évolutifs.

Nous présentons dans cet article une vision intégrée de ces diverses évolutions.

RÉSEAUX ET MARCHÉS VERTICAUX, UN CHANGEMENT DE PARADIGME

Les générations successives des réseaux mobiles ont historiquement poursuivi comme objectifs majeurs l'augmentation de la capacité, des débits et de la couverture, à des coûts rendant économiquement rentable leur déploiement. Avec l'avènement de la 5G, un changement de paradigme est introduit : la conception partiellement synergétique avec divers secteurs d'activité, tels que les transports et l'industrie du futur (mentionnons par exemple les initiatives 5G Automotive Association, 5G Alliance for Connected Industries and Automation et Software Republic).

De manière caricaturale, nous sommes passés d'une logique où les réseaux étaient conçus et déployés, puis le marché décidait « quoi en faire », à une logique dans laquelle un travail sur les possibles cas d'usage est réalisé très en amont et avec une participation directe de divers secteurs d'activité. Il en ressort des spécifications fonctionnelles et non fonctionnelles mieux adaptées aux besoins réels de ces secteurs.

Ainsi, les divers secteurs d'activité ne se limitent plus à être des utilisateurs des réseaux, mais deviennent des partenaires technologiques de ceux en charge des infrastructures numériques.

Les réseaux véhiculaires en tant qu'extension d'un réseau d'infrastructure représentent un exemple bien connu. Par ailleurs, les réseaux dits à l'*edge* (au bord) deviennent progressivement une partie intégrante des infrastructures du numérique.

Nous verrons par la suite que les interactions technologiques entre les réseaux et divers marchés verticaux (énergie, eaux, industries manufacturières, transports, bâtiment, etc.) se diversifient fortement avec les nouvelles générations de réseaux, donnant lieu à des évolutions majeures d'architectures, de services et de modèles économiques. Cela est encore plus significatif dans le cadre des réseaux privés ; ces derniers étant un des atouts clés pour la diffusion de la 5G et des réseaux du futur en général.

Cette interpénétration accrue se traduit par une dépendance fortement croissante des divers secteurs d'activité vis-à-vis des infrastructures du numérique, imposant sur celles-ci de nouvelles contraintes, notamment en termes de sûreté et de sécurité, mais également en termes de flexibilité (création automatique et dynamique de réseaux dans le cadre de systèmes virtualisés) et de latence (notamment dans le cadre du contrôle de sites industriels).

CONVERGENCE RÉSEAUX-CLOUD, MEC

En 2012, un groupe de 13 opérateurs télécom, dont Orange, ont publié un livre blanc¹ qui introduit le concept de virtualisation des fonctions réseau (NFV pour Network Functions Virtualization). Ce document a représenté en quelque sorte la validation de concepts qui étaient déjà étudiés, mais dont on doutait fortement de leur acceptabilité massive par ces acteurs.

Historiquement, les équipements de réseau télécommunication ont été construits comme des solutions *hardware*, spécifiques à chaque constructeur et en tant que boîtes noires. Cela présente de nombreux inconvénients : les difficultés et délais pour introduire de nouvelles fonctionnalités, le coût élevé et la dépendance technologique. Certes les réseaux sont standardisés, mais en fait seuls les interfaces entre grands blocs fonctionnels le sont.

Le document mentionné, partiellement motivé par le grand succès des centres de calculs appelés *data centers*, capables d'implémenter avec flexibilité une multitude de fonctionnalités, services et applications, préconise le remplacement des équipements dédiés par un *hardware* générique (serveurs informatiques) et l'implémentation *software* des fonctionnalités réseaux.

La *softwarisation* des réseaux, traitée en détail dans l'article d'Emmanuel Puig dans le présent numéro, est ainsi formalisée et fortement accélérée.

En conséquence, d'une part, de très nombreuses fonctions réseau sont aujourd'hui offertes en mode *cloud*, y compris pour la radio (*cloud-RAN*) et ce phénomène s'accélère. D'autre part, le *cloud* sort des grands *data centers* pour se disséminer jusqu'aux extrémités (le terme *edge* est utilisé dans la littérature) et notamment au niveau du MEC (*Mobile Edge Computing* ou *Multi-access Edge Computing*), permettant de nombreuses nouvelles applications, notamment celles sensibles à la latence ou nécessitant une protection particulière des données. Cela joue un rôle central dans les interactions avec les verticaux mentionnées plus haut. Le terme *edge* tel qu'utilisé dans ce contexte fait référence, soit à des points de présence des réseaux proches des clients finaux (bien plus proches que les grands *data centers*), ce qui est le cas du MEC, soit à des équipements côté usager, comme des terminaux, des voitures ou autres objets connectés.

¹ Network Functions Virtualization - An introduction, benefits, enablers, challenges & call for action (https://portal.etsi.org/NFV/NFV_White_Paper.pdf).

L'ensemble de ces évolutions prises conjointement nous font parler de convergence ou interpénétration réseau-*cloud*. L'utilisation du paradigme *cloud* dans les architectures de réseau permet à ces dernières d'introduire une très grande agilité et flexibilité. Le déploiement de ressources *cloud* dans les points de présence des réseaux permet quant à lui une grande diversité de nouvelles applications et usages.

Cette convergence facilite la mise en œuvre de diverses solutions réseaux, telles que le *slicing*, c'est-à-dire, la capacité de partitionner les réseaux en attribuant des parties (*slices*) à différents clients. Ce concept général ne diffère en principe pas trop du concept historique de réseau privé virtuel, sauf qu'ici la flexibilité est totale, chaque *slice* pouvant avoir des propriétés fonctionnelles et non fonctionnelles qui lui sont propres et avec la possibilité de faire évoluer dynamiquement et en temps réel les fonctions qu'il comporte et les ressources qu'il utilise.

Tout cela est possible grâce à la virtualisation des fonctions, qui de ce fait peuvent être ajoutées, éliminées ou déplacées dynamiquement. En effet, les fonctions, devenues purement logicielles et tournant sur du *hardware* générique, ne sont plus enfermées dans des boîtes noires.

Par ailleurs, les évolutions mentionnées plus haut sont un facilitateur majeur pour l'ouverture des réseaux. Cette ouverture se met en place de deux manières principales. Dans la première approche, les architectures des réseaux deviennent potentiellement plus ouvertes, permettant à une diversité d'acteurs de proposer des composants de réseau et aux opérateurs de construire leur réseau en déployant et interconnectant des composants de fournisseurs différents. C'est le cas notamment des solutions OpenRAN. Dans la deuxième approche, les réseaux proposent des interfaces, notamment des API, qui permettent à d'autres acteurs de contrôler une « partie » du réseau, typiquement celle qui leur est attribuée, notamment dans le cadre du *slicing*. Ce concept se généralise et introduit de nouveaux changements de paradigme, comme nous le verrons dans la section suivante.

Un élément clé des architectures ainsi définies est l'orchestrateur, l'intelligence qui permet à chaque instant d'identifier les fonctionnalités nécessaires, de les agencer et de leur attribuer les ressources dont elles ont besoin.

Toutes ces évolutions sont en cours, mais sont loin d'avoir abouti malgré leur intérêt certain. Une des raisons est que, même si la technologie répond plutôt bien aujourd'hui aux besoins de convergence mentionnés (dans des cadres restreints qui ne couvrent pas les évolutions à venir présentées plus bas), les filières industrielles concernées se trouvent dans une situation de concurrence de plus en plus forte. De ce fait, elles ont du mal à trouver les moyens et le temps pour opérer une plus forte convergence qui prenne en compte notamment de nouveaux modèles d'affaires ; nous y reviendrons plus loin dans cet article.

DES SERVICES RÉSEAUX AUX PLATEFORMES DU NUMÉRIQUE

L'étape suivante dans l'ouverture mentionnée plus haut consiste dans la possibilité pour des acteurs tiers de déployer leurs propres composants dans les infrastructures convergentes réseau-*cloud*. Ces fonctionnalités, développées par des tiers et déployées à travers d'APIs², portées ainsi par les infrastructures convergentes réseau-*cloud*, sont

² API : *Application Programming Interface*. L'idée principale d'une API est qu'elle permet à des acteurs externes d'interagir avec le système qui la propose sans savoir comment ce système est conçu et implémenté. Ces interactions peuvent être de multiples niveaux, tels que la demande d'une construction à la demande d'un réseau virtuel ou plus simplement, l'activation d'un service simple, par exemple de messagerie.

notamment en lien avec des services (de tout type, pas uniquement services réseau) et des applications fournies également par des acteurs tiers, pas nécessairement les mêmes que ceux qui déploient les fonctionnalités.

L'ouverture du cœur de réseau pour accueillir de telles fonctionnalités rend même possible l'émergence d'une place de marché de ces fonctionnalités déployées sur les infrastructures que nous traitons et ouvre ainsi de nombreuses opportunités et des risques importants.

L'idée générale, posée de manière « caricaturale », serait de voir les infrastructures convergentes réseau-*cloud* comme un PaaS (*Platform as a Service*) qui accueille de nouvelles fonctionnalités, de divers acteurs tiers, permettant de construire dynamiquement, à la demande, de nouveaux services et applications. Toute la chaîne, en partant du réseau lui-même, pouvant être construite à la demande et en temps réel.

Ce type de solutions est aujourd'hui plus prospectif, mais envisageable dans un horizon de 5 à 8 ans. Néanmoins, des verrous sont à lever, notamment dans la conception de nouveaux types d'orchestrateurs, dans la conception de mécanismes de validation des propriétés des composants acceptés dans la place de marché, dans la spécification des niveaux de qualité sur divers critères à exiger de ces composants, etc.

VERS LES SERVICES MULTISECTORIELS

Indépendamment des évolutions que nous venons de mentionner, une autre transformation se met en place. La numérisation, qui jusqu'à présent s'est faite en silos, par secteur d'activité, passe à une nouvelle phase, avec l'avènement de services multisectoriels. Dans ce cadre, des acteurs de secteurs d'activité différents collaborent pour développer conjointement des services fortement innovants s'appuyant sur les services et les données de chacun, tout en gardant la maîtrise de leurs données et infrastructures. Selon les modèles économiques choisis, ils peuvent aussi être en « coopération » (voir Figure 1). L'image suivante contient une représentation graphique de ces concepts.

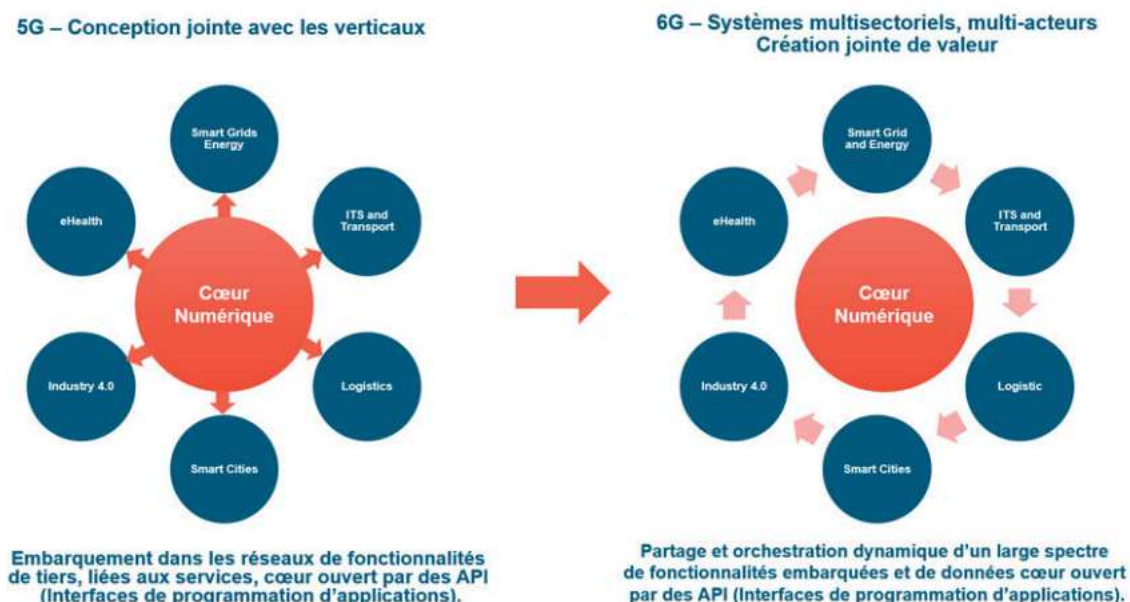


Figure 1 : Exemples de modèles économiques de coopération entre les acteurs.

La convergence réseau-*cloud* et l'ouverture des réseaux au travers des API mentionnées plus haut représentent clairement des facteurs d'accélération de cette tendance émergente.

Ici encore, les verrous sont multiples. Du point de vue technologique, l'orchestration dans un cadre multi-sectoriel pose de nouveaux problèmes. Pour les lecteurs familiers de l'architecture de l'Internet, on peut faire l'analogie avec l'introduction du routage externe, multi-systèmes autonomes. Notamment, les choix technologiques deviennent extrêmement dépendants des modèles d'affaires potentiels entre acteurs. À la différence du routage Internet, nous sommes ici face à une très grande diversité d'acteurs, de fonctionnalités à orchestrer, de contraintes à prendre en compte (notamment en lien avec l'accès aux données) et surtout à un ensemble d'acteurs et de fonctionnalités qui varie en permanence. La garantie de la qualité de service de bout en bout, le respect de bout en bout des politiques de sécurité et de protection des données de chaque acteur impliqué, le respect de leurs politiques en termes d'impact environnemental et de sobriété énergétique, ne représentent que quelques-uns des challenges à affronter.

Face à cette explosion de la complexité, de nouvelles approches de conception, de planification des opérations de contrôle et de maintenance, etc., basées sur de l'intelligence artificielle et l'utilisation de jumeaux numériques sont en cours d'étude, mais encore dans des stades très embryonnaires. Le lecteur intéressé est invité à lire notre autre article concernant ce sujet dans ce même numéro. De plus, une prise de conscience commence à apparaître concernant la surestimation du potentiel de l'IA et même de l'IA générative, et plus précisément de la capacité à trouver des modèles d'affaires permettant de faire face aux coûts engendrés par cette IA. Un article du 20 juin 2024 évalue que l'industrie de l'IA devrait générer un chiffre d'affaires annuel de 600 milliards de dollars pour être rentable³ et fait référence à une potentielle bulle autour de l'IA. Même si ceci concerne surtout la course au gigantisme des modèles généralistes, les IA génératives spécialisées pourraient être également impactées par un nouveau retournement de situation dans le financement de l'IA.

Nous sommes donc dans un cadre général qui permet de clairement identifier la voie, vu les avantages indiscutables des solutions décrites, mais dont il est très difficile d'évaluer les horizons de temps auxquels des solutions stables seront disponibles.

CONVERGENCE RÉSEAU-CLOUD, OUVERTURE ET *HYPERSCALERS*, UN FUTUR INCERTAIN

L'ensemble des évolutions mentionnées dans les paragraphes précédents pose de manière encore plus centrale la question des investissements et du partage de la valeur entre les filières historiques et notamment entre opérateurs de télécommunications et *hyperscalers*⁴. Depuis plus de 20 ans (Google, par exemple, a été créé en 1999 et le premier iPhone date de 2007), cette question du partage de la valeur est soulevée par les opérateurs qui sont les principaux investisseurs dans les infrastructures de réseaux qui supportent les services des *hyperscalers*. Ils disposent ainsi du contrôle de ces réseaux et des ressources confortables de communication associées. Les évolutions mentionnées risquent de réduire fortement le positionnement clé des opérateurs dans la chaîne de valeur. Si nous ajoutons à cela les investissements des *hyperscalers* dans des câbles sous-marins de très grande capacité, la croissance forte du marché des réseaux non terrestres (avec notamment les constellations de satellites en orbites basses et moyennes) avec la concurrence accrue qui se dessine dans ce domaine et la multiplicité d'opérateurs de télécommunications en Europe qui se traduit par le fait que les entreprises en question ont un poids financier limité (en comparaison aux autres acteurs mentionnés), les risques pour les opérateurs

³ <https://www.sequoiacap.com/article/ais-600b-question>

⁴ On appelle *hyperscalers* les quelques compagnies américaines qui dominent le marché mondial du *cloud*.

de télécom, mais aussi pour les équipementiers fournisseurs historiques de ces derniers, sont évidents.

Mitiger ces risques impose une vision holistique des acteurs impliqués, des évolutions technologiques, des nouveaux modèles d'affaires, de la réglementation, etc. Les décisions sont difficiles : la *softwarisation* des réseaux a des avantages notables pour les opérateurs, mais elle les rend plus vulnérables aux positionnements d'autres acteurs. De même, l'ouverture des réseaux, à la mode OpenRAN, présente un intérêt pour les opérateurs, mais met en risque le positionnement des constructeurs, qui néanmoins n'ont pas le choix et doivent suivre. L'IA semble *a priori* un moyen clé pour la maîtrise de la croissance très rapide de la complexité, mais elle a ses propres challenges et un risque de ralentissement dans les investissements. Le multi-sectoriel est une voie de forte création de valeur, très probablement la prochaine phase de la transformation numérique, mais la disposition des acteurs pour y avancer n'est pas claire et les modèles économiques pour accélérer les processus sont à définir.

Dans ce cadre, de nouvelles formes d'organisation de l'écosystème pourront éventuellement émerger, ouvrant la porte à des rôles nouveaux, facilitateurs des relations technologiques ou économiques entre les parties prenantes.

Tout ceci accroît de manière notoire les enjeux autour de la souveraineté numérique, voire de la souveraineté tout court, étant donnée l'imbrication croissante du numérique avec tous les secteurs d'activité. Une réflexion de fond, globale, multipartite semble s'imposer dans l'intérêt de tous à moyen terme et cela même si la pertinence et les modalités ne semblent pas évidentes à court terme.

Le *cloud* dans la chaîne de valeur industrielle du numérique

Par **Quentin ADAM**
CEO de Clever Cloud

À l'ère de la transformation numérique, la France et l'Europe occidentale sont confrontées à une révolution économique majeure. Les technologies numériques, et en particulier le *cloud computing*, transforment profondément notre manière de créer et de distribuer la richesse. Cette transformation numérique représente une opportunité sans précédent, mais également un défi de taille pour la France, qui doit jongler entre innovation technologique et souveraineté numérique. Malheureusement, ce sujet ne reçoit pas toujours l'attention nécessaire, et les impacts importants de cette révolution sont souvent minimisés.

Le *cloud computing*, devenu un pilier central de la transformation numérique, bouleverse le paysage économique mondial. En facilitant l'accès aux services numériques et en augmentant l'efficacité opérationnelle des entreprises, il redessine les contours de l'économie mondiale. Cependant, cette révolution soulève également des questions fondamentales sur notre souveraineté économique et notre dépendance vis-à-vis des géants technologiques mondiaux.

Alors que la France, aux côtés de ses partenaires européens, cherche à tirer parti de cette transformation, il est crucial de comprendre les dynamiques économiques qui en découlent et de développer des stratégies robustes pour garantir une participation équitable et bénéfique à l'économie numérique mondiale. Plus important encore, il est essentiel de préserver notre modèle social et notre mode de vie, actuellement précarisés et menacés à moyen terme par notre faible empreinte industrielle dans ce secteur.

Cet article se propose de revisiter les bases, de dresser un état des lieux, et de formuler des moyens d'action pour privilégier une réaction pertinente face à cette problématique. Et pour commencer, comment est-on devenu riche ?

COMPRENDRE LA SOURCE DE LA RICHESSE DE L'EUROPE OCCIDENTALE

La richesse de l'Europe occidentale s'est construite au fil du temps en passant d'une économie de flux à une économie d'*assets*, jouant un rôle crucial dans l'évolution des systèmes économiques modernes.

L'économie de flux se caractérise par la prise d'une fraction de la valeur générée par le travail au moment où il est effectué. Ce modèle repose sur une circulation continue et rapide des biens, des services et des capitaux. Dans une telle économie, la valeur ajoutée est souvent réalisée par l'intermédiaire d'échanges, de transactions et de la distribution de produits.

Bien que dynamique, l'économie de flux ne favorise pas nécessairement l'accumulation de richesse à long terme. C'est ici qu'intervient l'économie d'*assets*. Contrairement à une simple économie de flux, une économie d'*assets* repose sur la détention et l'accumulation de biens matériels et d'outils de production. Ces actifs permettent non seulement la création de valeur, mais aussi sa capitalisation et sa transmission à travers le temps.

L'industrialisation a été le catalyseur qui a transformé les actifs en sources de capitalisation significatives. Avec l'avènement de l'industrialisation, les machines, les usines et les technologies sont devenues des leviers puissants pour accroître la productivité. Si l'on schématise, l'apparition du moteur, puis des automates (comme la machine de Watt, considérée comme le premier régulateur PID de l'histoire), a permis de remplacer les êtres humains par des machines, capturant ainsi la marge produite par ces machines pour rémunérer le capital. Cela a transformé les matières premières en produits finis, augmentant la valeur ajoutée et facilitant l'accumulation de richesses, tout en augmentant la part de revenu capitalistique des détenteurs des outils de production.

Cette accumulation a rendu possible, et parfois nécessaire, la financiarisation des actifs. Elle a permis aux entreprises de lever des capitaux, d'investir dans des outils de production ou de nouvelles technologies, et de s'étendre au-delà de leurs marchés locaux tout en augmentant leurs marges. La capacité de capitaliser sur ces actifs a conduit à la création de grandes entreprises industrielles qui ont dominé l'économie mondiale.

Il est essentiel de comprendre que la clé du développement économique européen a reposé sur l'industrialisation : remplacer les humains (utilisés principalement pour leur force physique) pour produire à bas coûts des produits complexes, tout en capitalisant sur cette richesse. Cela a déplacé la répartition entre capital et travail, générant l'opportunité de financiariser les outils de production.

LE MOTEUR DE LA PRESTATION INTELLECTUELLE : L'INDUSTRIALISATION DES CERVEAUX

L'essor de l'industrialisation manufacturière a joué un rôle clé dans la libération de la main-d'œuvre de nombreuses tâches physiques, ouvrant ainsi la voie à l'expansion du secteur tertiaire au cours du dernier quart du XX^e siècle. Toutefois, avec l'avènement de l'ordinateur et des technologies numériques, un nouveau paradigme a émergé : l'industrialisation des cerveaux. De la même manière que le moteur a remplacé le travail physique de l'humain, l'ordinateur commence à remplacer les tâches cognitives, entraînant des transformations économiques et sociologiques majeures.

Historiquement, des penseurs tels que Karl Marx et Anne-Robert-Jacques (Baron de) Turgot ont analysé la dynamique des rendements décroissants du capital dans des économies traditionnelles, allant de l'agriculture à l'immobilier, en passant par l'industrie. Ils ont observé que dans une économie classique, chaque unité supplémentaire de capital produit un retour décroissant, en raison de contraintes physiques et de l'épuisement des ressources, ce qui signifie que les bénéfices marginaux tendent à diminuer avec l'augmentation de l'investissement en capital.

Ce scénario classique dans une économie industrielle traditionnelle montre bien que, contrairement à ce qui se passe dans le secteur numérique, l'expansion et l'investissement continus dans des conditions physiques et logistiques limitées mènent inévitablement à une réduction de l'efficacité du capital investi. On peut le transposer dans un modèle d'immobilier locatif, et ainsi de suite...

En contraste, dans l'économie numérique, des économistes contemporains comme Brian Arthur ont souligné que nous assistons à des rendements croissants du capital. En raison de l'absence de coût de copie et de l'élimination des modèles d'usure des outils de

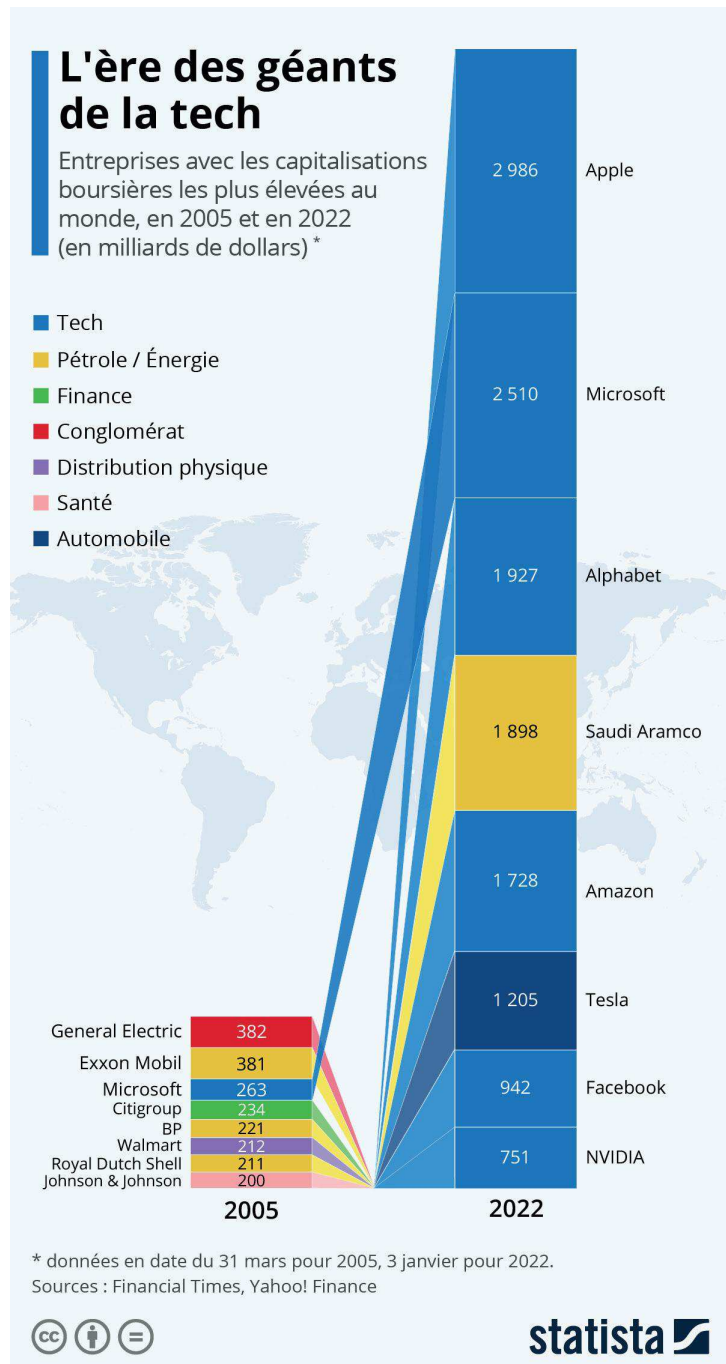
production intellectuelle, un investissement en capital dans le numérique peut générer des retours exponentiellement croissants, notamment grâce aux effets de réseau et aux économies d'échelle. Par exemple, une plateforme en ligne telle que Facebook ou Google peut ajouter des millions d'utilisateurs sans que les coûts n'augmentent significativement, amplifiant considérablement sa rentabilité. De façon ironique, on pourrait dire que l'économie digitale est en train de vicier le capitalisme dans son fonctionnement.

Figure 1 : Entreprises avec les capitalisations boursières les plus élevées au monde, en 2005 et en 2021, en milliards de dollars (Sources : Financial Times et Yahoo Finance).

Cette évolution est illustrée par les changements dans les valorisations des plus grandes entreprises mondiales entre 2005 et 2020. En 2005, les entreprises les plus valorisées étaient diversifiées à travers des secteurs comme l'énergie, l'industrie et la finance. En 2020, le paysage a radicalement changé avec une prédominance des entreprises technologiques telles qu'Apple, Amazon et Microsoft. Cette concentration dans le secteur numérique démontre comment le rendement croissant du capital a permis à ces entreprises de dominer le marché mondial. Leur valorisation est devenue massivement supérieure à celle des plus grandes capitalisations quinze ans auparavant.

C'est pourquoi je qualifie cette ère de révolution industrielle de la prestation intellectuelle et soutiens qu'il est

aujourd'hui essentiel de posséder les « machines-outils » de cette nouvelle révolution : les logiciels et les bases de données associées. De plus, avec l'émergence de l'intelligence artificielle (IA), nous observons une accélération de ce phénomène, rendant encore plus critique la maîtrise de ces technologies pour assurer la compétitivité dans l'économie numérique.



QU'EST-CE QUE L'IA ?

L'histoire des sciences est jalonnée par des débats entre les approches réductionnistes, popularisées par René Descartes, et les théories de la complexité, développées plus tard par des penseurs comme Edgard Morin. Descartes a posé les fondements du réductionnisme scientifique, qui préconise de diviser les problèmes complexes en leurs éléments les plus simples pour en faciliter l'étude. Cette méthode a dominé la science pendant des siècles, permettant d'innombrables découvertes en décomposant les systèmes en parties compréhensibles.

Toutefois, Edgard Morin a critiqué cette approche pour sa tendance à négliger les interactions et les interdépendances qui caractérisent les systèmes complexes. Il a plaidé pour une méthode qui reconnaît et embrasse la complexité, en particulier dans les sciences humaines et sociales, où les phénomènes ne peuvent pas toujours être résolus par la division et l'isolement.

Ainsi, la complexité est une approche de philosophie des sciences qui a des applications multiples, telles qu'en sciences sociales ou en médecine avec la biologie des systèmes. Elle est particulièrement pertinente en intelligence artificielle, comme le montre l'exemple de la vision humaine. L'IA a permis des progrès significatifs en adoptant une approche d'apprentissage similaire à celle des humains. En utilisant des réseaux de neurones profonds, l'IA apprend à reconnaître et à interpréter des images sans comprendre pleinement le processus sous-jacent. Cette méthode d'apprentissage par l'expérience permet à l'IA de capter les subtilités de la vision humaine, démontrant ainsi la puissance de l'apprentissage inspiré par la complexité.

L'intelligence artificielle moderne repose sur plusieurs piliers fondamentaux :

- **Données** : Essentielles pour l'apprentissage et la prédiction, les données alimentent les algorithmes d'IA, leur permettant de reconnaître des modèles et d'effectuer des tâches avec une précision croissante. Il faut des données bien rangées et triées, telles que l'index d'un moteur de recherche ou le contenu utilisateur des réseaux sociaux.
- **Algorithmes** : Les algorithmes, notamment ceux du *machine learning* et des réseaux de neurones, sont au cœur de l'IA, permettant le traitement complexe des données. Il faut de bons mathématiciens pour les développer, par exemple ceux de l'ENS, dont quasiment toutes les équipes du monde comptent un ancien élève.
- **Puissance de calcul** : L'accès à des processeurs toujours plus performants, comme les GPU, a permis des avancées significatives en permettant des calculs rapides et complexes. En clair : le *cloud computing*.
- **Intégration technologique** : L'efficacité de l'IA dépend aussi de son intégration dans des systèmes existants, permettant des automatisations et des interactions sophistiquées dans divers domaines.

Oui, dans cette liste, on retrouve de nombreux combats abandonnés par notre pays, et c'est pour cela qu'il faut absolument revoir notre politique.

LA RÉVOLUTION NUMÉRIQUE ET LE *CLOUD COMPUTING*

Le *cloud computing* ne se limite pas à la virtualisation des machines ou au simple stockage de données. Il représente une approche intégrée de fourniture de matériel, de services, de logiciels et d'énergie à un niveau de service unifié et à la demande. Cette définition élargie montre que le *cloud* est une plateforme complexe qui va bien au-delà des fonctions informatiques traditionnelles.

Les clés du *cloud* sont souvent méconnues et se basent sur plusieurs piliers essentiels :

- Unification de la facturation et affectation de la facturation : Le *cloud* offre un système unifié de facturation qui simplifie la gestion des coûts pour les entreprises. Cette transparence permet aux organisations de comprendre précisément où leurs ressources sont dépensées et d'optimiser leurs dépenses informatiques.
- *Marketplace* et taille du catalogue : À l'image des grossistes, le *cloud* propose une vaste place de marché où les utilisateurs peuvent accéder à un large éventail de produits et services. Cette diversité permet aux entreprises de choisir les solutions qui répondent le mieux à leurs besoins spécifiques, tout en profitant de la flexibilité et de l'évolutivité offertes par le *cloud*.
- Unification de l'accès et de l'expérience utilisateur : Le *cloud* intègre l'accès aux services et l'expérience utilisateur en un seul endroit, offrant ainsi une interface cohérente et simplifiée. Cela réduit la complexité pour les utilisateurs finaux et améliore l'efficacité opérationnelle.
- Gestion des identités et des accès (IAM) et auditabilité : Le *cloud* met en place des systèmes robustes de gestion des identités et des accès, assurant que seuls les utilisateurs autorisés peuvent accéder aux ressources. De plus, l'auditabilité intégrée permet de suivre et de vérifier les activités des utilisateurs, renforçant ainsi la sécurité et la conformité.

Les produits insérés dans ce schéma évoluent vers une plus grande abstraction et automatisation, mais il est crucial de ne pas confondre ces produits transitoires avec les clés du *cloud* qui façonnent l'avenir de l'usage informatique. Alors que le *cloud* continue de se développer, il est important de comprendre ces éléments fondamentaux pour saisir son impact profond sur les entreprises et les infrastructures numériques.

Les géants du *cloud*, tels qu'Amazon Web Services (AWS), Microsoft Azure et Google Cloud, sont avant tout des entreprises de logiciels, tant sur le plan historique que financier. Si le *cloud* était principalement une question d'infrastructure, des opérateurs traditionnels comme Verizon ou AOL seraient leaders du marché. Cela se reflète dans leurs tarifs : AWS propose une gamme de services avec des prix souvent jusqu'à huit fois supérieurs à ceux d'OVH, par exemple. On peut se demander si Amazon, initialement un acteur du *retail*, est si inefficace dans ses achats d'infrastructure qu'il doit vendre à des prix aussi élevés pour maintenir sa marge, ou bien si ce qui est réellement facturé est le logiciel.

Il est crucial de comprendre que l'infrastructure dans le *cloud* est une conséquence de la vente de logiciels, et non l'inverse. Cette perspective doit guider notre stratégie : plutôt que de s'appuyer sur des opérateurs téléphoniques pour résoudre nos défis dans le *cloud*, il est impératif de stimuler l'innovation dans les services logiciels, et donc chez les éditeurs. La France et l'Europe doivent se concentrer sur le développement de logiciels qui définissent l'avenir du *cloud*, en investissant dans la recherche et le développement pour créer des solutions capables de rivaliser avec les géants du marché.

Un aspect essentiel de la R&D dans le *cloud* est la fusion entre matériel et logiciel, une approche qui a démontré son efficacité dans le succès de produits comme l'iPhone ou les Mac Mx chez Apple. Cette stratégie est également visible dans des projets comme Nitro et Graviton chez AWS, les TPU (Tensor Processing Units) chez Google, et les *switches* Sonic chez Microsoft. En optimisant le logiciel et le matériel ensemble comme un seul produit industriel intégré, ces entreprises parviennent à augmenter considérablement leurs marges et à simplifier la gestion opérationnelle.

En Europe, une alliance formée par des entreprises telles que Clever Cloud, Vates, Kalray, 2CRSI, SiPearl, et ZML dans divers projets de recherche vise à livrer ce type de

résultats. Leur objectif est de maximiser la rentabilité financière tout en réalisant des économies d'énergie et de consommation massive, s'inscrivant dans une vision écologique. Ces collaborations montrent que le *cloud* doit être compris comme un produit de service industriel, où l'intégration du matériel et du logiciel est cruciale pour créer des solutions efficaces et durables.

Des projets tels que S3NS et Bleu¹ illustrent les défis des partenariats internationaux déséquilibrés. Lorsque les entreprises françaises gèrent uniquement l'infrastructure (rendement décroissant du capital) tandis que le développement logiciel (rendement croissant du capital) est contrôlé par des partenaires étrangers, cela crée un déséquilibre. Cela peut mener à une dépendance et une perte de contrôle économique. Ces modèles sont économiquement insoutenables et doivent être évités. Pour maximiser leur autonomie technologique et économique, les acteurs français doivent rechercher des alliances qui répartissent équitablement les bénéfices et renforcent la compétitivité nationale. Même si cet article n'aborde pas les implications juridiques et stratégiques, ces considérations sont essentielles pour préserver les intérêts nationaux.

UNE COLONIE NUMÉRIQUE

L'Europe, et en particulier la France, est devenue une « colonie numérique », comme l'indique le rapport d'information n°443 (2012-2013) du Sénat, déposé le 20 mars 2013 par Catherine Morin-Desailly. La souveraineté technologique de notre pays est effectivement compromise. Les géants du logiciel étrangers contrôlent non seulement l'infrastructure numérique, mais aussi les données personnelles et professionnelles. Ce contrôle s'apparente à un pillage de nos richesses, comme le montrent des affaires d'intelligence économique telles que celles d'Alstom ou de Technip, détaillées dans les travaux de Frédéric Pierucci.

Nous vivons un cycle d'accaparement de nos ressources patrimoniales, ce qui met en péril notre autonomie stratégique et menace notre modèle économique. Depuis 2005, dernière année où la France a affiché une balance commerciale positive, le déficit commercial s'est accru chaque année.

Ce déficit a des répercussions directes sur notre capacité à financer les services publics et notre modèle de retraites. Notre incapacité à devenir un acteur majeur de la révolution industrielle en cours entrave notre développement économique et met en danger notre modèle social. Il est impératif que la France et l'Europe développent des infrastructures numériques souveraines pour reprendre le contrôle de leurs données et réduire cette dépendance. C'est le rôle de chaque citoyen, dans sa position, de mettre en œuvre une politique volontariste à son échelle pour inverser ce processus.

...CAR NE REMONTANT PAS LA CHAÎNE DE VALEUR INDUSTRIELLE

Dans le domaine du *cloud computing*, le logiciel et la gestion des systèmes d'information (DSI) suivent un modèle de chaîne de valeur industrielle, similaire à celui utilisé dans la fabrication et la production, avec des fournisseurs classés par rangs différents. Pour créer un environnement sain, il est crucial d'appliquer la méthodologie de sélection de fournisseurs et de gestion du marché héritée de l'apprentissage industriel. Ce modèle nécessite une gestion attentive des chaînes d'approvisionnement, en diversifiant les fournisseurs

¹ Offres de « *cloud* de confiance » Bleu (les technologies de Microsoft proposées par Orange et Capgemini) et S3NS (celles de Google avec Thales).

pour éviter la dépendance excessive à un seul acteur. Il est stupéfiant de voir combien d'entreprises communiquent fièrement sur leur capacité à s'être placées en dépendance d'un fournisseur, qui, le plus souvent, finira par devenir un concurrent. Ce manque de clairvoyance est préoccupant.

L'affaire du rachat de VMware par EMC illustre bien les dangers de cette dépendance. Lorsque le numérique favorise la généralisation de monopoles en raison de l'absence de limites physiques à la copie, il appartient aux acheteurs de prendre des décisions stratégiques pour diversifier leurs partenaires technologiques et maintenir la concurrence, limitant ainsi le comportement moutonnier de nombreuses structures. Les entreprises doivent être vigilantes dans la gestion de leur marché pour éviter de devenir captives de quelques géants du *cloud*, et ainsi conserver leur liberté d'action et d'innovation.

Par ailleurs, il est essentiel de bannir l'idée de « vagues technologiques successives » que l'on pourrait manquer ou rattraper, comme un surfeur californien qui saute d'une vague à l'autre. Utiliser la métaphore de l'immeuble haussmannien pour illustrer cela est pertinent : construire un immeuble solide nécessite de solides fondations et une structure stable (notre chaîne de valeur industrielle). Dire que les fondations ne sont pas stables, que le toit fuit, mais que la peinture des toilettes du deuxième étage est magnifique et que cela rattrape l'immeuble serait absurde. C'est pourtant cette approche qui prévaut lorsque l'on se précipite vers de nouveaux marchés sans consolider les fondamentaux, en admettant que l'on aurait « loupé cette vague ».

L'exemple de l'intelligence artificielle, mentionné plus haut, est criant : les ingrédients de la chaîne de valeurs sont ceux qui sont délaissés au profit de la nouveauté qu'elle représente, alors qu'ils sont essentiels à son déploiement. L'IA promet de révolutionner de nombreux secteurs, mais elle doit s'appuyer sur des infrastructures solides et des données fiables. Ce n'est pas la « vague technologique » du moment qui garantit le succès, mais la capacité à construire sur des bases solides et à intégrer ces innovations de manière durable.

Il est impératif que les acteurs du numérique français, et en particulier les ingénieurs, relisent leurs cours d'industrialisation et les appliquent : le logiciel n'est que le *firmware* d'une machine-outil industrielle, et ce n'est pas un monde virtuel à la Matrix. Les mêmes principes économiques doivent gouverner sa gestion.

PERSPECTIVES POUR LA FRANCE

La France est reconnue pour l'excellence de sa formation scientifique et technologique, avec des institutions prestigieuses telles que les École Normale Supérieure (ENS), l'X et les Grandes Écoles du numérique. De nombreux ingénieurs français formés dans ces établissements contribuent aux succès internationaux de l'intelligence artificielle et occupent des postes clés dans les entreprises technologiques de premier plan à travers le monde. Notre système éducatif, largement financé par les fonds publics, forme des ingénieurs de haut niveau, mais ces talents finissent souvent par travailler pour des entreprises étrangères, fournissant ainsi de la rentabilité et de l'innovation à des concurrents plutôt qu'à l'économie française.

Même si ces talents sont à la base de la création de *start-up* en IA, cette fuite des cerveaux est problématique car elle reflète une gestion inefficace des ressources humaines en France. Les ingénieurs français manquent souvent de perspectives attrayantes sur le marché local, car les décisions de gestion technologique sont prises de manière inappropriée. Il est crucial que les élites technologiques, ayant une compréhension approfondie des enjeux et des innovations, soient impliquées dans la direction des sujets technologiques.

Un regard sur les fondateurs et dirigeants des GAFAM, tels que Steve Jobs, Elon Musk, Bill Gates, Eric Schmidt, Larry Page, Sergey Brin, Mark Zuckerberg, et Steve Wozniak, montre que ceux-ci sont toujours des ingénieurs et des visionnaires technologiques qui comprennent intrinsèquement les produits et services qu'ils développent. Disons les choses clairement : dans le top mondial des entreprises technologiques, 100 % des fondateurs et gestionnaires sont des ingénieurs, sans exception. Si vous croyez le contraire, je vous invite à relire la biographie de chacun d'eux pour vous en convaincre. Leur succès réside dans leur capacité à combiner une expertise technique avec une vision stratégique. En France, moins de 15 % des fondateurs des entreprises du French Tech Next 120 sont des ingénieurs, et on s'étonne de ne pas faire émerger de GAFAM français ?

Pour stimuler l'innovation et renforcer la compétitivité, il est essentiel que les élites technologiques prennent part aux décisions stratégiques et dirigent les projets innovants, en créant un écosystème où les compétences humaines sont valorisées et cultivées. Cela enclencherait une mutation rapide de l'écosystème.

Malgré les défis posés par la domination des géants du *cloud*, l'espoir n'est pas perdu pour la France. Une stratégie efficace consiste à soutenir l'innovation locale par la commande publique plutôt que par la simple subvention. Il est crucial de rappeler que tous les GAFAM ont bénéficié de commandes publiques massives dans leur développement, notamment à travers des initiatives telles que la DARPA aux États-Unis. Cette stratégie doit inspirer des initiatives en Europe, comme JEDI, qui mérite d'être encouragée et soutenue.

La diversification des approvisionnements au sein des entreprises françaises est également essentielle pour cultiver un marché dynamique. Les *leaders* locaux doivent se concentrer sur le soutien au marché des éditeurs (économie d'*assets*) plutôt que sur les sociétés de services numériques (ESN) qui opèrent selon une économie de flux. En privilégiant les éditeurs, les entreprises peuvent renforcer la valeur et la durabilité de leurs investissements technologiques.

En ce qui concerne l'*open source*, il est important d'adopter une approche axée sur la création de communs numériques déployés par des éditeurs, et non sur l'idée que l'*open source* équivaut à la gratuité. L'*open source* doit permettre d'accélérer le développement du marché par le biais de communs et de renforcer les compétences domestiques et la gouvernance européenne. Il doit favoriser les éditeurs locaux plutôt que de servir de complément aux revenus des ESN pour réduire les coûts. En d'autres termes, l'*open source* doit être un modèle de partage collaboratif de la R&D pour augmenter la vitesse et non un moyen de paupérisation des éditeurs locaux.

Quant au *soft power*, il est essentiel d'adopter une attitude de non-réponse face aux sollicitations américaines, car le *lobbying* permanent représente une force majeure. Par exemple, la seule contrepartie du plan Marshall fut l'ouverture du marché des salles de cinémas en France. Rappelons-nous que, bien que les États-Unis soient des alliés, ils sont également des concurrents dans le monde économique, et que l'expression "America First" est porteuse dans une large part de leur population. Il est important de naviguer dans cette compétition avec une stratégie qui protège et promeut nos propres intérêts.

LA LOI DES RENDEMENTS ACCÉLÉRÉS

La loi des rendements accélérés, popularisée par Ray Kurzweil dans son livre *The Singularity is Near*, stipule que le progrès technologique s'accélère de manière exponentielle. Cela signifie que chaque avancée rend possible des innovations encore plus rapides, réduisant les délais et les coûts associés.

Par exemple, il fallait une armée de concepteurs pour réaliser des effets spéciaux médiocres en 3D dans les années 1980, alors qu'aujourd'hui, quelques personnes suffisent grâce à des suites logicielles très abouties, sans même parler des progrès récents amenés par l'IA.

La loi des rendements accélérés démontre sans peine que, contre le défaitisme ambiant, les niveaux d'investissement exorbitants colportés par les géants américains du *cloud* ne sont pas nécessaires. Par ailleurs, on peut débattre de leurs calculs d'investissement qui incluent des éléments tels que le passage de câbles optiques intercontinentaux, ce qui me semble hors spectre et juste bon à faire grossir la facture du *lobbying* pur et simple dans un seul objectif : nous décourager.

Et pourquoi ? Car d'ici à 2030, le marché du *cloud* devrait croître d'environ 20 % par an². Nous sommes donc à un stade où le marché est un "blue ocean", libre et facile à conquérir, et la façon la plus simple de gagner une bataille est d'être le seul sur le terrain. Le soft power américain vise à nous convaincre de ne pas aller sur le champ de bataille, de ne même pas chercher à concourir, alors que nous avons toutes les technologies en Europe pour le faire, comme le montre le travail d'inventaire d'Euclidia³. Il ne faut pas déclarer perdue une bataille qui n'a pas été menée.

Les petites et moyennes entreprises (PME) jouent un rôle crucial dans l'innovation disruptive, souvent en étant à l'avant-garde des nouvelles technologies et des modèles d'affaires novateurs. Ce phénomène est bien illustré par le concept de destruction créatrice de Joseph Schumpeter, qui souligne comment l'innovation issue des petites structures peut renverser les marchés établis et créer de nouvelles opportunités économiques. En France, les PME représentent une source majeure de créativité et de dynamisme économique.

Pour maximiser leur impact, il est essentiel que les grandes entreprises apprennent à collaborer efficacement avec ces petites structures. Les grandes organisations peuvent bénéficier de l'agilité et de la capacité d'innovation rapide des PME, tandis que les petites entreprises peuvent tirer parti des ressources et des réseaux des grandes entreprises pour accélérer leur développement.

Un cadre législatif et financier favorable à l'innovation est également crucial. L'instauration d'un *Small Business Act* à la française pourrait offrir des incitations fiscales et des subventions ciblées pour les *start-ups* et les PME, stimulant ainsi la création d'emplois et l'innovation. Cela permettrait de faciliter l'expérimentation et la prise de risques, éléments clés pour que les petites entreprises puissent réaliser leur potentiel disruptif.

En favorisant un écosystème où les grandes et petites entreprises collaborent étroitement, la France peut renforcer son tissu économique et se positionner comme un leader dans l'innovation numérique à l'échelle mondiale. Cela nécessite une volonté de part et d'autre de reconnaître la valeur que chaque type de structure peut apporter à l'autre et d'intégrer cette collaboration dans leurs stratégies d'innovation.

CONCLUSION : VERS UN INTERNET HUMANISTE

À l'origine, l'internet, ou « inter-réseaux », n'était pas unique et centralisé, mais plutôt un ensemble de liens entre des réseaux eux-mêmes décentralisés. Au fil du temps, et sous l'influence des contraintes géopolitiques et culturelles, ces liens se sont distendus, donnant naissance à des îlots de visions distinctes de l'internet. Il est illusoire de penser qu'internet est unitaire dans le monde car sa réglementation ne l'est pas.

² <https://www.alliedmarketresearch.com/cloud-services-market>

³ <https://www.cloudrepo.eu/>

L'internet des régimes illibéraux, illustré par l'exemple de l'internet chinois ou russe, est caractérisé par un contrôle étroit exercé par l'État, avec une censure et une surveillance omniprésentes. Ce modèle renforce la souveraineté numérique nationale mais souvent au détriment de la liberté d'expression et de la transparence, cloisonnant ainsi les idées et la population. Par exemple, en Chine, pour publier sur internet, il faut un numéro de publication qui permet un contrôle strict par le pouvoir des éditeurs.

L'internet de l'économie de l'attention et de la spoliation des données personnelles est dominé par les entreprises américaines, qui exploitent massivement les données personnelles pour capter l'attention des utilisateurs et maximiser les profits. Ce modèle, centré sur la monétisation des données, soulève des préoccupations croissantes concernant la vie privée et la concentration du pouvoir entre les mains de quelques géants technologiques.

L'internet humaniste, centré sur la vision européenne, cherche à préserver la vie privée et les droits fondamentaux des citoyens grâce à une réglementation stricte, comme le Règlement général sur la protection des données (RGPD). Cette vision est de plus en plus suivie à travers le monde. Des cultures millénaires et riches, comme le Japon, ont très tôt adopté une approche similaire et compatible. Aujourd'hui, des lois comme la loi 25 au Québec et les mesures restrictives de la CNIL nigérienne montrent une alliance possible entre des valeurs humanistes portées par des pays réfléchis, combinant une culture du nord globale et des pays émergents qui ne veulent pas reproduire les vices du marché subis par les Européens et qui mettent en danger leur modèle social. Il y a une alliance objective à construire.

Pour la France et l'Europe, l'enjeu est de promouvoir un internet humaniste qui allie protection des droits, innovation technologique et compétitivité économique. Cela nécessite une implication active dans la définition des normes internationales, en s'appuyant sur les valeurs de transparence, de sécurité, et de respect des droits fondamentaux.

En cultivant cette vision humaniste, la France peut non seulement protéger ses citoyens, mais aussi servir de modèle pour un internet ouvert qui concilie innovation et éthique. C'est en défendant ces valeurs que nous pourrons garantir un avenir numérique respectueux de la dignité et de la liberté de chacun.

Mais c'est aussi, et surtout, le seul chemin pour assurer un avenir à notre style de vie en redressant notre balance commerciale. Le risque d'effondrement n'est pas à prendre à la légère, nous sommes en danger. Seule la mobilisation efficace de chaque membre de la communauté nationale, à son niveau, permettra de sortir de l'ornière et de reprendre notre déploiement. C'est une responsabilité collective que nous devons assumer, en abandonnant immédiatement le fatalisme et les œillères face aux écueils.

De la manufacture à l'infrastructure : des conséquences pour la régulation

Par Pierre-Jean BENGHOZI

Directeur de recherche émérite au CNRS et à l'École polytechnique

En quelques décennies, le numérique s'est imposé comme une composante essentielle de nos sociétés. La régulation des infrastructures numériques, telles que les réseaux de télécommunications et les plateformes de données, représente un enjeu stratégique majeur. Elles nécessitent des investissements massifs et soulèvent des questions complexes de souveraineté, d'accès équitable et de répartition de la valeur. La régulation doit également s'adapter aux évolutions rapides des technologies et des marchés, tout en gérant les risques systémiques liés à la cybersécurité et à la fiabilité des réseaux. Cet article examine ces défis et propose des pistes pour une régulation efficace et équitable.

En quelques décennies, le numérique s'est imposé comme une composante essentielle de nos sociétés, transformant profondément les économies, les cultures et les interactions sociales. Les infrastructures de connectivité sont désormais aussi indispensables que celles de l'eau ou de l'électricité. Maîtriser ces infrastructures est crucial pour la souveraineté et l'accès aux biens communs publics, que ce soit pour les télécommunications, le stockage ou les données. Cela nécessite des investissements massifs en R&D et en déploiement, favorisant souvent la concentration au profit des acteurs et pays les plus puissants.

La régulation de ces infrastructures numériques représente dès lors un enjeu stratégique majeur en termes de souveraineté, d'équité et de partage de la valeur. Pour offrir à tous l'accès à des réseaux de communications performants, les nations doivent mobiliser plusieurs leviers. Un premier levier est le calibrage et les modalités des investissements pour couvrir tout un territoire et donc le partage des efforts entre public et privé. Un deuxième tient à l'articulation entre soutien aux infrastructures, aux équipements et aux usages : entre disponibilité et appropriation. Il s'agit de concilier le dimensionnement des tuyaux avec la croissance des usages, c'est-à-dire aussi de répartir contributions et revenus entre opérateurs qui déploient et plateformes applicatives.

DE LA MANUFACTURE À L'INFRASTRUCTURE, UN MONDE QUI CHANGE¹

Poids des grandes plateformes de l'Internet, fiscalité internationale, difficulté de définir une politique industrielle et une souveraineté économique... Les acteurs politiques, les

¹ Le texte développe certaines idées présentées dans BENGHOZI P.-J. (2022), « Savoir passer d'une économie de la manufacture à l'économie des infrastructures », in *Informatisation et Entreprises*, pp. 17-20, Institut de l'Économie éd.

observateurs de la société et les économistes ont du mal à appréhender ces mouvements et à suggérer des réponses. Car nous ne sommes plus dans le modèle de la manufacture sur laquelle Smith ou Ricardo avait construit les principes séminaux de l'économie. Nous sommes désormais dans une économie de l'infrastructure : les puces, les applications, les réseaux de télécommunication et d'énergie, les plateformes et serveurs de données. Ces composants constituent les ressources partagées pour tous les services numériques.

Par des modes d'appropriation flexibles et des possibilités de dématérialisation, ils ont redéfini les rapports aux territoires et aux organisations industrielles. En permettant un suivi plus efficace des processus de production et des pratiques de consommation, ils ont contribué ainsi au renouvellement des modèles d'affaires. Le poids et l'évolution des infrastructures numériques en fait donc des *"game changer"*. Au-delà de l'émergence des seules grandes plateformes applicatives, elles offrent à de nouveaux acteurs l'opportunité d'entrer sur des marchés à la fois comme fournisseurs d'équipement, opérateurs de réseaux, parties prenantes d'applications, fournisseurs de services, intermédiaires de données... Pensons, par exemple, aux transformations induites avec le développement de la fibre optique, des constellations de satellites ou de la 5G.

UNE ÉCONOMIE DE COÛTS FIXES : « JE SAIS BIEN, MAIS QUAND MÊME »

L'économie des infrastructures est caractérisée par de hauts coûts fixes et de faibles coûts marginaux de production, à même de modifier profondément les modalités habituelles de la concurrence.

D'une part, le développement des infrastructures numériques requiert des investissements massifs en recherche et développement (R&D) ainsi qu'en déploiement de technologies. Ces investissements, souvent concentrés entre les mains de quelques acteurs puissants, peuvent favoriser la concentration du pouvoir économique et technologique.

Mais, d'autre part, le couplage entre un très haut niveau de coûts irrécupérables et des coûts marginaux quasi nuls rend particulièrement sensible la concurrence par les prix. Certaines raisons sont valables pour toute infrastructure : les investissements matériels sont difficiles à redéployer une fois qu'ils sont installés. D'autres sont plus spécifiques au numérique : les technologies sont évolutives et les investissements sont en partie fongibles². En outre, les composants techniques (protocoles, langages, terminaux) et les logiciels peuvent être combinés pour concevoir des services de toute nature à partir d'un même réseau. Les conséquences sont un rythme très rapide d'innovation et un environnement intensément concurrentiel.

Les dynamiques portées par des stratégies numériques d'intégration verticale conduisent à remettre en cause l'appréhension des infrastructures de télécommunications comme reposant sur plusieurs couches techniques autonomes (composants, équipements, protocoles, interfaces, architecture, applications, terminaux), au profit de perspectives réglementaires intégrant des couches techniques poreuses et de moins en moins séparables et à la base des innovations émergentes de rupture. La prise en compte de couches statique apparaît décalée à l'heure d'une convergence dynamique et d'innovations naissant dans un cadre où les concurrences s'opèrent d'une couche à l'autre. Certains gros secteurs utilisateurs (aéronautique, jeux vidéo ou culture par exemple) ont eu un rôle déterminant dans la structuration des grands acteurs du numérique (respectivement avec Sabre pour les réseaux, Nvidia pour les puces, Akamai pour les réseaux de diffusion de contenu

² L'architecture et les pilones des stations de réseaux mobiles s'adaptent ainsi d'une génération de fréquences à l'autre.

ou en anglais *content delivery network* – CDN). Les acteurs dominants eux-mêmes se concurrencent à partir de cœurs de métiers ancrés à différents niveaux³ et ils investissent lourdement chacun dans les infrastructures de réseaux (fibre, câbles sous-marins, satellites, serveurs).

Cette évolution appelle à remettre en cause des perspectives traditionnelles. Les infrastructures sont habituellement vues comme plus structurantes que les applications car elles appelleraient des investissements et une capitalisation plus importants. Or, les schémas actuels (pensons à l'IA) montrent que les applications peuvent être plus intensives en investissement que les réseaux : de l'ordre de dizaines de milliards d'euros pour les réseaux fixes en France, de centaines de milliards d'euros pour le développement des IA.

DES CONSÉQUENCES ÉCONOMIQUES ET DE RÉGULATION

Ces développements interrogent sur les modalités de répartition de la valeur et des investissements entre biens et services, tuyaux et contenus, ouvrant la voie à des stratégies de convergence et de redéfinition des marchés où les enjeux d'interopérabilité sont déterminants.

En effet, la multiplicité des utilisateurs et parties prenantes suscite un goulot d'étranglement majeur : le manque – ou à l'inverse le foisonnement – des normes et standards. D'un côté les acteurs dominants ont la tentation oligopolistique de constituer des infrastructures propriétaires constituant autant d'écosystèmes hétérogènes en concurrence pour « capturer » les consommateurs. De l'autre, des innovations d'infrastructure peuvent contribuer à la fragmentation des réseaux. On l'observe dans les accès radio mobile de la 5G (RAN⁴) qui se démultiplient avec l'évolution du partage des liaisons terrestres, le déliement des réseaux RAN et des réseaux centraux, l'apparition d'opérateurs virtuels, la concurrence en terminaison avec le wifi...

Dans de tels cadres, la perspective historique des économistes et des régulateurs a été de considérer les infrastructures comme un monopole naturel et « une facilité essentielle » qu'il serait absurde de dupliquer (une liaison de chemin de fer par exemple). Les enjeux de régulation sont alors celui du contrôle d'une forme de monopole, le niveau de tarification du marché de gros et la soutenabilité des modèles économiques associés, la neutralité du réseau, l'incitation à l'innovation, la sécurité et la fiabilité. Pourtant, en matière de télécommunications, la tentation des régulateurs a longtemps été de favoriser une concurrence par les infrastructures. Ainsi, loin d'envisager les infrastructures de télécommunications comme un « monopole naturel » le régulateur français⁵ a très tôt considéré nécessaire, pour l'animation du marché, que des concurrents développent leurs propres infrastructures jusqu'aux limites économiquement raisonnables. Dans de tels cadres de modèles propriétaires intégrés et non interopérables, les enjeux de régulation sont alors surtout l'interopérabilité, l'accent sur les possibilités de sortie (*opt out*) et de basculement sur un autre opérateur (*switching*) du consommateur, la tarification, la mutualisation des investissements d'infrastructure et la prévention des cartels.

La difficulté tient à ce que les infrastructures évoluent techniquement à la fois par des ruptures technologiques rapides et imprévisibles (pensons à l'émergence des constel-

³ Microsoft (système d'exploitation) vs. Amazon (webservice) vs. Apple (constructeur de matériel).

⁴ Un Radio Access Network est le composant du réseau cellulaire permettant l'interface entre les terminaux et le cœur du réseau.

⁵ En phase d'ailleurs avec les autorités européennes.

lations satellitaires) mais aussi très progressivement. Ainsi, historiquement, la fibre a été déployée graduellement par les opérateurs en commençant par les grands réseaux dorsaux internationaux, puis dans les réseaux de transport longue distance nationaux. Elle a ensuite été mise en œuvre dans les réseaux de collecte et ce mouvement se poursuit aujourd’hui, avec le FttH⁶, au-delà du nœud de raccordement, dans la boucle locale, jusqu’à l’abonné final.

De telles évolutions sont inéluctables et il s’agit donc de savoir encadrer la transition d’une génération d’infrastructure à l’autre. Mais les rythmes d’évolution des technologies d’infrastructure, de leurs déploiements et de leurs utilisations sont très différents. Ainsi, la durée de vie attendue est de 50 ans pour la fibre et de l’ordre de 10 ans pour la 5G ou 6G. Il faut cependant d’abord construire une infrastructure avant de s’en servir et les processus opérationnels de déploiement prennent des années⁷. Dans le même temps, le rythme des évolutions des couches applicatives est beaucoup plus rapide, de l’ordre de mois ou d’années : l’apparition et la large diffusion des IA génératives (ChatGPT, Copilot ou Gemini) se sont faites en à peine 1 an. Les difficultés de concevoir des infrastructures pour le long terme tiennent à ces rythmes très différents ouvrant le risque de gaspillages (calibrage ou technologie du réseau devenant inutile par rapport à des nouveaux besoins) et la difficulté de mettre en phase déploiements et usages émergents.

Dans ces environnements technologiques en constante évolution, les régulateurs doivent naviguer entre intervention *ex ante* et *ex post*. La régulation *ex ante* fige à l’avance les structures industrielles des réseaux (prix et marché) en fixant aux acteurs les « règles du jeu ». La régulation *ex post* évite de réguler trop tôt pour ne pas geler le marché et donner un avantage au “first mover” en l’absence de certitudes sur les choix techniques à assurer... mais elle conduit à intervenir trop tard lorsqu’il s’agit de marchés fortement évolutifs où existent de forts écarts entre les pouvoirs de marché des acteurs. Cela a été le cas avec la régulation des grandes plateformes du numérique.

INFRASTRUCTURES ET RISQUES SYSTÉMIQUES

Au-delà des enjeux techniques et économiques, de nouveaux registres de la régulation touchent désormais les infrastructures, ceux qui concernent leurs impacts environnementaux ainsi que la gestion des risques systémiques aux effets particulièrement lourds, l’un comme l’autre. Qu’ils proviennent de mauvaises conceptions, de défauts techniques, de cyber-attaques ou d’emballement des usages, les défaillances des infrastructures peuvent toucher les transports, les services financiers, les réseaux électriques ou de télécommunications et sont alors susceptibles de causer des dommages dévastateurs à tous les pays. Les organisations modernes sont en effet extrêmement dépendantes à l’égard de la communication. Leurs évolutions rapides et fréquentes disqualifient en permanence les modes de sécurisation préexistants et les régulateurs doivent mettre en place des cadres robustes pour protéger les infrastructures critiques et garantir la résilience des systèmes numériques.

Dans une infrastructure ouverte à différents acteurs et reposant, par nature, sur des interconnexions, chaque intervenant et chaque composant technique risque de contribuer à l’effondrement et la dislocation complète du fonctionnement, exacerbant, par un effet « domino » certains impacts bien au-delà de leurs effets initiaux. La virtualisation et les réseaux sans fil, le poids de l’algorithmie dans leur gestion, la mobilité et la disponibilité

⁶ Fiber to the Home.

⁷ Voir plus d’une décennie pour la complétude d’un réseau fixe très haut débit, BENGHOZI P.-J. (2023), « Infrastructures numériques et aménagement du territoire : impacts économiques et sociaux du Plan France très haut débit », France Stratégie, janvier, 205 pages.

permanente des connexions créent des vulnérabilités inédites. Celles-ci sont d'autant plus difficiles à gérer quand elles s'inscrivent dans des séquences d'événements tels que des catastrophes naturelles, entraînant des ruptures lourdes des infrastructures, conduisant à la saturation des réseaux de communication puis leur déconnexion.

Comme l'a récemment montré, à l'été 2024, la panne de Microsoft et Crowstrike, les problèmes de sécurité se diffusent rapidement à travers le monde et ne se confinent plus à un pays donné car une part importante des communications traverse les frontières sans que l'utilisateur final ne s'en rende même compte. Quand les infrastructures de télécommunications relevaient de monopoles d'État, la sécurité restait maîtrisable : ces grandes organisations publiques contrôlaient leur réseau, elles constituaient un club limité d'opérateurs se connaissant et ayant l'habitude de coopérer. Avec l'ouverture des marchés, les infrastructures sont maintenant détenus et gérés par de multiples acteurs, souvent en concurrence entre eux et donc moins enclins à collaborer.

CONCLUSION

Le développement des infrastructures a toujours été un motif important de réglementation gouvernementale mais elle est difficile à mettre en œuvre avec l'augmentation des interconnexions et l'implication d'un nombre plus important d'acteurs. Cette complexité constitue un vrai enjeu car le régulateur – acteur public – est légitime quand son objectif est clair, compris et partagé. Or il doit désormais poursuivre simultanément plusieurs séries d'objectifs parfois opposés⁸ : utilité des réseaux pour les consommateurs, services innovants favorisant services, usages et activités, opérateurs performants, soutien à l'économie nationale, mise en valeur de territoires isolés... Cette variété des ambitions suppose de maîtriser plusieurs registres de compétences (économique, financier, juridique, technique, opérationnel, connaissance des consommateurs et usages). Elle appelle alors, de fait, à partager la responsabilité entre plusieurs intervenants. Le Gouvernement arrête le cadre légal et fixe les orientations de politique publique ; la Commission européenne détermine un cadre réglementaire et assure supervision et contrôle ; les autorités administratives sectorielles (Arcep, Arcom, ART, CRE...) régulent les infrastructures spécifiques et animent leur marché ; les autorités ou institutions généralistes (CNIL, AdIC, ANSSI...) encadrent différents aspects (données, concurrence, cybersécurité) de manière transversale ; le Conseil d'État et les tribunaux arbitrent et valident les décisions⁹.

Le nouveau cadre de la régulation est donc celui d'une régulation multi-acteurs reposant sur l'action croisée de ces différentes instances. C'est aussi une régulation multi-niveaux dans la mesure où les cadres international et européen se superposent désormais systématiquement au cadre national.

Comme le dit un adage juridique bien connu, toute régulation est, *prima facie*, territoriale. La légitimité des États dérive en effet de la souveraineté sur un territoire physique. Pendant longtemps, il y avait de ce fait correspondance entre les autorités politiques de l'espace physique et la régulation des infrastructures de communications fixe ou mobile. Le numérique a progressivement bouleversé ce système fondé sur les frontières physiques. Les communications et les transactions s'opèrent entre des entités sans lien

⁸ Le code des postes et des communications électroniques (CPCE), en liste 14 différents dans son article L.32-1.

⁹ Il est d'ailleurs intéressant de noter que face à la convergence entre télécommunications, audiovisuel et informatique, les différents pays ont fait des choix différents en matière de structuration : en Grande-Bretagne, l'OFCOM regroupe par exemple tuyaux et contenus. En France même, les décideurs publics ont eu régulièrement la tentation de regrouper et de fédérer les différentes autorités sectorielles.

nécessairement avec la localisation géographique des autres parties. La virtualisation des réseaux rend possible de transférer des services dématérialisés dans des pays au cadre fiscal ou réglementaire plus favorable et les grandes entreprises technologiques peuvent alors exploiter les disparités entre les pays et poser des défis à la souveraineté économique des nations.

Au niveau international, les aspects techniques et normatifs se discutent dans des organisations parfois très anciennes¹⁰, mais aussi, désormais, dans des organisations interétatiques non gouvernementales associant les différentes parties prenantes (recherche, États, industriels, société civile) selon des modes inédits¹¹. Au niveau européen, l'intervention de la Commission est désormais déterminante dans la définition des orientations et du cadre d'intervention (« paquets » Télécoms), en précisant les modalités de décentralisation et subsidiarité (cadre général fixé par les directives), en assurant des actions de contrôle (recommandations et cadre d'harmonisation *ex ante*, *veto ex post*, sanctions), en arbitrant les incohérences entre régulations nationales (selon les spécificités techniques, géographiques, historiques, économiques, politiques), en instituant¹² une coopération institutionnalisée des régulateurs européens.

Doivent ainsi s'opérer des interrelations juridiques inédites entre développement territorial des infrastructures d'un côté, internationalisation des acteurs et des couches techniques de communication et de services de l'autre côté. C'est tout le dilemme face auquel se trouve aujourd'hui la Commission européenne : elle développe d'un côté tout un ensemble de directives visant à réguler puissamment les plateformes dominantes (DSA, DMA, Data Act, IA Act), tout en confinant de fait les opérateurs de télécommunications à leurs marchés nationaux en refusant les démarches de consolidation continentales à même de contrebalancer la puissance des grandes plateformes du numérique.

L'harmonisation des différents systèmes juridiques est pourtant d'autant plus importante que sous l'effet des mutations en cours, ces systèmes ont connu des vagues successives d'évolution, en fonction des pays et régions de l'Internet : harmonisation au niveau continental en Europe, extraterritorialité pour les États-Unis, cantonnement pour la Chine ou la Russie. Si ces mouvements ne semblent pas nécessairement converger, la percolation de certaines orientations (RGPD en matière de données personnelles, antitrust pour les magasins d'applications) montrent cependant que les décisions prises dans différents États peuvent aussi se rapprocher. Pour prendre en charge les impacts de l'internet sur l'économie mondiale et la globalisation des échanges, les instances non techniques de gouvernance internationale (UE, G8, G20...) ont d'ailleurs été amenées à aborder des sujets qui relevaient surtout, jusque-là, des seuls cadres de régulation sectoriels.

¹⁰ L'Union Internationale des Télécommunications (UIT) a été fondée en 1865 et la Conférence européenne des administrations des postes et télécommunications (CEPT) en 1959.

¹¹ World Wide Web Consortium (W3C) fondé en 1994 ou The Internet Corporation for Assigned Names and Numbers (ICANN) en 1998.

¹² À la main des régulateurs ou de la Commission.

Convergence des infrastructures numériques : un point de vue économique

Par Laurent BENZONI

Université Paris-Panthéon-Assas et Tera Consultants

L'activité des opérateurs de télécommunications (TELCOS) est territoriale, celle des fournisseurs majeurs de contenus et d'applications (GAFAM) est mondiale. Il s'ensuit pour les TELCOS une taille plus réduite, une dette plus élevée, une moindre rentabilité. Avec la convergence, les GAFAM descendent la chaîne de valeur de l'infrastructure numérique pour offrir des services intégrés et capter une valeur plus grande. Ils se rapprochent des TELCOS mais restent à l'écart de la boucle locale. Chaque abonné d'une TELCO requiert un investissement considérable en comparaison des services d'un GAFAM.

Si les TELCOS ont échoué à remonter vers l'amont de la chaîne de valeur, l'émergence du *slicing* offre l'opportunité d'offrir une connectivité différenciée pour mieux valoriser les conditions d'accès à leurs réseaux et à leurs abonnés. Cependant, le respect de la neutralité du net et le recours à l'intelligence artificielle pour implémenter ces offres limiteront les marges de manœuvre des TELCOS.

Le progrès technique constant enregistré en amont dans la conception et la fabrication des composants électroniques se traduit en aval par un effondrement des coûts de traitement, de stockage et de transport des signaux numérisés. Désormais, les réseaux de télécommunications numérisés assurent une connectivité favorisant l'intégration/convergence des réseaux avec les systèmes de traitement et de stockage des informations qui se situaient historiquement à leur périphérie (*edge*). S'est ainsi formé un écosystème se traduisant par des offres de services nouvelles et innovantes que ce soit pour les ménages (vidéo à la demande par abonnement par exemple) ou les entreprises (*software*, plateforme, infrastructure, *network as a service*).

Cet écosystème fonctionne grâce à la complémentarité technique de tous les éléments qui le composent. Désormais et *de facto*, les réseaux de communication constituent l'un de ces éléments au sein d'un ensemble bien plus vaste qualifié d'infrastructure numérique, spécifiant ainsi la convergence entre des secteurs et des activités qui fonctionnaient auparavant de façon relativement indépendante. La synergie croissante et forte des activités pour l'offre de services au niveau des consommateurs, ménages ou entreprises, brouille, voire efface, les frontières existantes entre différents métiers. Les acteurs issus de ces différents métiers doivent désormais se positionner dans l'écosystème de l'infrastructure numérique en fonction des risques et des opportunités que constitue ce nouvel environnement en évaluant au mieux leurs forces et leurs faiblesses.

Cet article apporte données et faits pour alimenter concrètement la réflexion autour de ce phénomène de convergence qui, de l'avis consensuel des analystes et observateurs, constituera un enjeu majeur pour l'Union européenne et notamment, pour la France.

CONVERGENCE : LE POIDS ÉCONOMIQUE DÉCROISSANT DES RÉSEAUX DE COMMUNICATIONS ÉLECTRONIQUES EN DÉPIT D'UNE FONCTION TOUJOURS ESSENTIELLE

Historiquement, les réseaux constituaient le socle de l'infrastructure des communications. En 2010, l'Arcep établit que le montant du chiffre d'affaires des opérateurs de réseaux fixes de communications électroniques atteignait 16,5 milliards d'euros tandis que celui des mobiles montait à 19,5 milliards, soit un total de 36 milliards d'euros dégagés par les opérateurs de réseaux de communication.

Treize années plus tard, ces deux marchés ne comptent plus que pour 32 milliards, une baisse de 11 % en euros courants, soit bien plus encore en euros constants (cf. Tableau 1). La baisse du chiffre d'affaires dans les services mobiles qui paraît paradoxale au regard de sa diffusion spectaculaire dans la société s'explique avant tout par la guerre des prix déclenchée par l'arrivée d'un quatrième opérateur.

Tableau 1 : Le marché des services de télécommunication fixes et mobiles en France de 2010 à 2023 en Mds € (Source : Observatoire des marchés des communications électroniques, Arcep).

	2010	2023	Variation
Fixes	16 544 €	16 998 €	+ 3 %
Mobiles	19 458 €	15 125 €	- 22 %
Total	36 002 €	32 123 €	- 11 %

L'explosion de la consommation de données et la révolution numérique ne se lisent donc pas dans les ventes des opérateurs de réseau installés en France. Elles se retrouvent en revanche, bien plus dans les chiffres de leurs investissements qui ont dépassé 150 milliards d'euros sur la période 2010-2023, soit plus de 11 milliards d'euros par an en moyenne. Il faut noter que ce total inclut 10 milliards d'euros versés à l'État par les opérateurs pour l'obtention de leurs autorisations d'utilisation des fréquences pour les services mobiles¹.

Au-delà des activités de réseau, l'Arcep collecte aussi le chiffre d'affaires des services à valeur ajoutée relevant de ses prérogatives et les ventes de terminaux mobiles des opérateurs, soit un total supplémentaire de 4,4 milliards d'euros en 2023.

L'infrastructure numérique dépasse désormais amplement le seul périmètre des réseaux et des services sous contrôle de l'Arcep. Malheureusement, il n'existe pas de représentation standardisée de l'écosystème. Les périmètres des marchés et des activités fluctuent en fonction des informations disponibles ou des besoins des études. Deux découpages verticaux seront retenus ici :

- Un premier découpage représente la chaîne de valeur de l'écosystème numérique articulée autour de 7 secteurs d'activités partant de la production de contenus en amont jusqu'à la distribution des biens et services en aval (cf. en marron sur la Figure 1). Ce découpage est utile pour repérer le positionnement des acteurs, en particulier ceux intégrés sur plusieurs maillons de la chaîne de valeur. Par exemple, Amazon est présent sur les maillons de la production de contenus, de la fourniture de contenus et d'applications, d'*enabler*, de fourniture de terminaux et de la distribution.

¹ Ce chiffre n'inclut pas le 1 % du chiffre d'affaires versé pour l'usage des fréquences, soit 2,8 milliards d'euros versés en sus à l'État par les opérateurs sur la période.

- Un second découpage représente l'infrastructure numérique articulée autour de 6 fonctions nécessaires pour assurer la bonne délivrance des services produits en amont aux consommateurs qui les utilisent en aval (en bleu sur la Figure 1 notamment). Utilisé notamment par l'Arcep pour appréhender l'évolution de l'interconnexion entre les réseaux, ce second découpage constitue, comme montré sur la Figure 1, une subdivision des quatre maillons centraux de la chaîne de valeur présentés dans le premier découpage.

Dans la partie grisée entre les deux découpages sont présentés synthétiquement 24 segments afin de concrétiser et mieux spécifier les services et métiers concernés à chaque étape de la chaîne de valeur.

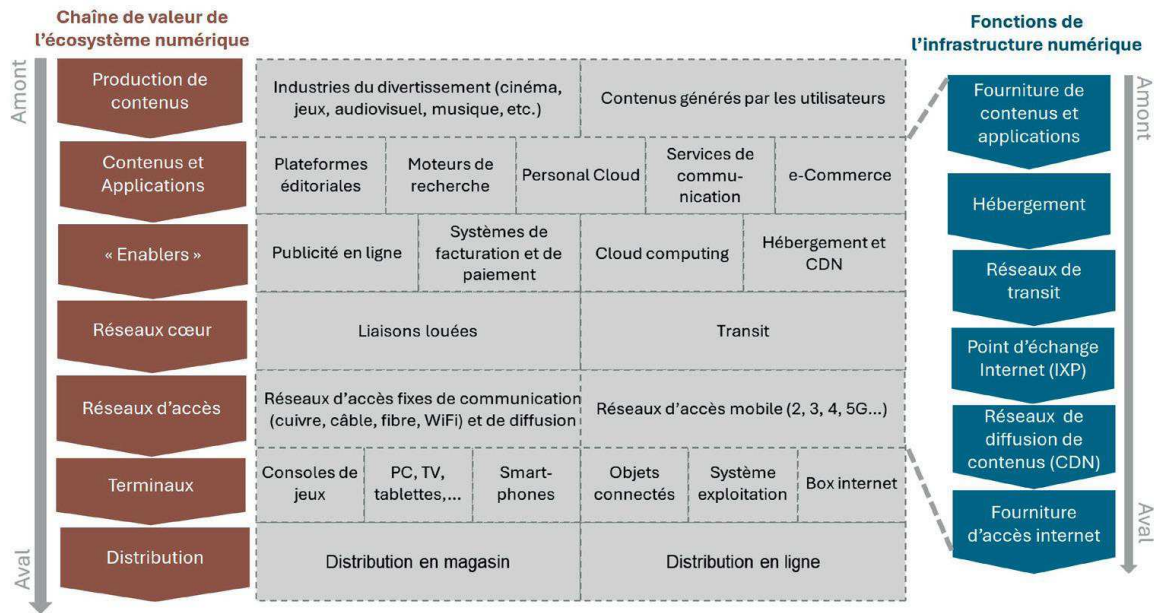


Figure 1 : Chaîne de valeur de l'écosystème numérique et fonctions de l'infrastructure numérique (Source : d'après l'Arcep et Tera Consultants).

Faute de collecte d'informations centralisée et normalisée, l'évaluation économique précise des composantes constitutives de la chaîne de valeur ou des fonctions de l'infrastructure numérique reste très complexe. Les chiffres publiés sont peu cohérents entre eux et d'une fiabilité incertaine. On peut procéder par touches afin de cerner les grandeurs en cause.

Au niveau mondial, une étude de 2022 pour la GSMA (Global System for Mobile Association) indiquait que 60 % du marché des infrastructures numériques était le fait des fournisseurs de contenus et d'applications (FCA), contre 15 % pour les fournisseurs d'accès Internet (FAI) que sont principalement les opérateurs de réseaux de télécommunications (cf. Figure 2). Une étude de 2010 évaluait la part des FAI à près de 50 %. Leur poids

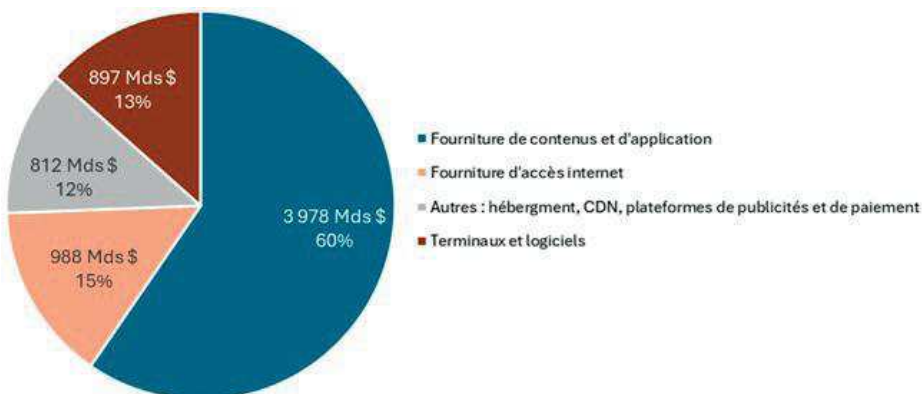


Figure 2 : Chiffres d'affaires mondiaux des fonctions de l'infrastructure numérique (Source : GSMA).

économique dans le total de l'infrastructure numérique aura donc été divisé par trois en un peu plus de 10 ans.

Au niveau national, le marché du *cloud computing* était estimé à 21 milliards d'euros en 2023 tandis que le marché de l'e-pub dépassait 9 milliards d'euros, celui du e-commerce atteignait 160 milliards, la vidéo à la demande par abonnement pointait à plus de 2 milliards, etc.

Sans surprise, la France connaît une dynamique identique à celle observée au niveau mondial : les réseaux de communications électroniques qui assurent l'indispensable connectivité pour la délivrance des services au niveau des ménages et des entreprises voient leur part économique se réduire très rapidement au bénéfice des fonctions situées en amont.

Les flux de trafic à l'interconnexion quantifient ce phénomène. Les flux entrants dans les réseaux des quatre principaux FAI qui émanent quasi-uniquement des FCA sont dix fois plus importants que les flux sortants émis par les clients des FAI en France.

Il appert que la dynamique de la croissance de l'écosystème est impulsée par l'amont avec une forte création de valeur par des services constamment renouvelés et innovants.

Cependant, techniquement, tous les maillons de la chaîne de valeur et les fonctions de l'infrastructure numériques sont solidaires entre eux. La qualité d'un service délivré aux consommateurs dépendra de la qualité de son maillon le plus faible.

La disponibilité pour les consommateurs d'un accès au réseau de haute qualité, fiable et sécurisé, disponible sur l'ensemble du territoire, constitue toujours le goulet d'étranglement de toute la chaîne de valeur située en amont. Les services peuvent présenter une connectivité différenciée en fonction des éléments de l'infrastructure qui ont été mobilisés pour les produire, les stocker et les transporter. Mais la qualité finale de tout service numérique, quel qu'il soit, reste totalement contingente de la qualité de la connectivité proposée par l'accès aux consommateurs finaux.

Les réseaux de communications électroniques continuent ainsi d'assurer cette fonction particulièrement critique pour toute la chaîne de valeur située en amont de ces réseaux même si leur poids économique au sein cette chaîne s'est considérablement réduit au cours des dernières années.

Économiquement, la solidarité entre les différentes fonctions de l'infrastructure numérique stipule une complémentarité qui implique de rechercher une répartition équitable des charges et des revenus entre les parties prenantes afin que le jeu coopératif soit gagnant-gagnant pour être stable.

CONVERGENCE : LE JEU COMPLEXE DU POSITIONNEMENT SUR LA CHAÎNE DE VALEUR DE LA VALEUR DE L'INFRASTRUCTURE NUMÉRIQUE

Au niveau des utilisateurs finaux, la convergence n'évoque pas les problématiques techniques d'interpénétration des métiers et des fonctions de l'infrastructure numérique. Pour les utilisateurs, la convergence s'exprime concrètement par des offres de services « clés en main » leur évitant la nécessité d'acquérir et d'assembler par eux-mêmes différentes applications, services, fonctions ou composantes de la chaîne de valeur pour satisfaire leurs besoins. Elle se traduit ainsi par un nombre d'interlocuteurs qui se réduit au fur et à mesure que les offreurs intègrent les fonctions de la chaîne de valeur de l'infrastructure numérique.

Ainsi pour les ménages, « convergence » signifie qu'un seul abonnement permet de cumuler l'accès fixe haut ou très haut débit, à un ou plusieurs accès mobiles avec ou sans terminaux inclus (*smartphone*, voire TV), l'accès à des services de vidéo à la demande, à la diffusion de la musique, à lecture de la presse, à la sauvegarde des données personnelles, à la domotique, etc.

Pour les entreprises, la convergence se traduit par la multiplication d'offres de type SaaS, PaaS, IaaS ou NaaS (cf. Figure 3).

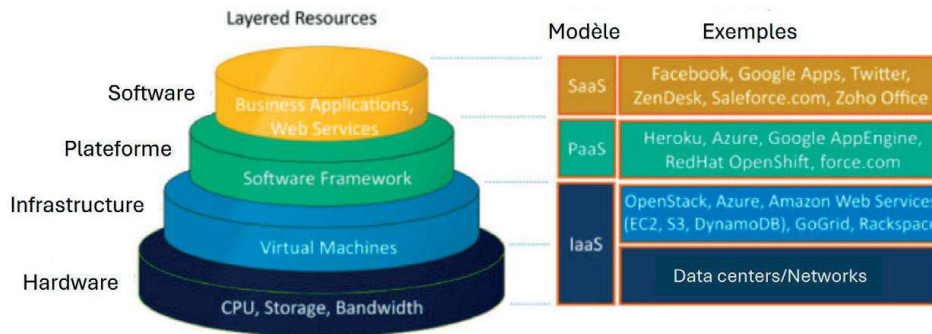


Figure 3 : SaaS, Paas, IaaS/Naas (Source : d'après "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges", Alcardo Alex Barakabitze, Arslan Ahmad, Rashid Mijumbi, Andrew Hines).

Ces offres se généralisent grâce à l'intégration croissante des fonctions de l'infrastructure numérique aboutissant à une forme de recentralisation externalisée des systèmes d'information : le *cloud computing*. Celui-ci est dominé au niveau mondial par Amazon Web Services (AWS), Microsoft Azure et Google Cloud Platform. Selon l'Autorité de la concurrence, ces trois acteurs qualifiés d'*hyperscalers* captent respectivement 46, 17 et 17 % des dépenses en infrastructures et applications de *cloud* public en France.

Les acteurs de l'offre étaient présents sur un ou plusieurs maillons de la chaîne de valeur, jamais sur tous. Pour proposer des offres de convergence, les acteurs disposent alors de trois options, non exclusives l'une de l'autre, pour combler leurs « trous » au sein de la chaîne de valeur :

- acheter les prestations manquantes aux acteurs des autres maillons ;
- créer des partenariats avec d'autres acteurs positionnés sur des maillons complémentaires pour bâtir des offres communes ou concurrentes ;
- internaliser les fonctions des maillons manquants pour capter toute la valeur économique du maillon et, partant, de la convergence.

Il s'ensuit une évolution structurelle des positionnements sur la chaîne de valeur (cf. Figure 4) : les FCA descendent vers l'aval (flèche 1 sur la Figure 4) pour se situer au plus près des utilisateurs avec, pour les plus gros FCA, une présence directe chez les utilisateurs à travers les terminaux (Microsoft sur les PC et les consoles de jeux, Google avec le système d'exploitation Android, Apple avec les *iphones*, *ipads*, Amazon avec la liseuse Kindle, Meta avec les casques VR, etc.). Pour ces grands FCA, seule la boucle locale des réseaux de télécommunications échappe jusqu'à présent à leur stratégie d'intégration sur la totalité de la chaîne de la valeur du numérique et sur les fonctions de l'infrastructure numérique.

À l'opposé les FAI remontent vers l'amont en se positionnant dans le transit, l'hébergement, dans les plateformes de paiement, de publicité, éditoriale voire plus amont encore dans la production de contenus avec néanmoins le peu de succès que l'on sait, à l'instar d'Orange avec OCS ou plus encore d'ATT avec Time Warner (flèche 2 sur la Figure 4).

Les acteurs des CDN se placent en aval des serveurs dans les réseaux des FAI (CDN internes ou *on-net* CDN en anglais) et intègrent en amont le transit dans leurs offres (flèche 3 sur la Figure 4). Les acteurs du transit descendent vers l'aval pour proposer la fonction de CDN pour se situer ainsi plus près des FAI afin d'accroître leur attractivité auprès des FCA (flèche 4 sur la Figure 4).

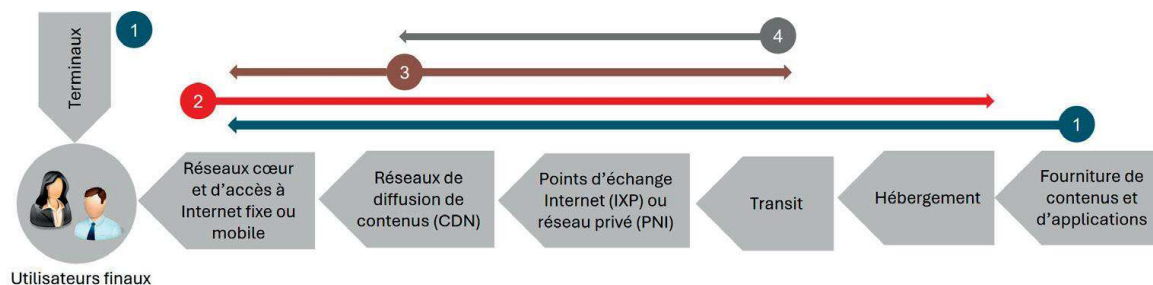


Figure 4 : Le repositionnement des acteurs sur l'infrastructure numérique
(Source : Tera Consultants et Arcep).

CONVERGENCE : TELCOS VERSUS GAFAM

Comme le montre la Figure 4 ci-dessus les acteurs les plus actifs de la convergence sont les opérateurs de télécommunications qui remontent la chaîne de valeur et les acteurs dominant la fourniture de contenus et d'applications qui descendent vers l'aval de cette chaîne. Ces derniers étant ce qu'il est convenu d'appeler les GAFAM (Google-Apple-Facebook-Amazon-Microsoft).

Une comparaison de quelques fondamentaux économiques de ces deux catégories d'acteurs permet d'appréhender les sous-jacents des modèles d'affaires et d'en inférer leurs forces et faiblesses respectives.

Le Tableau 2 ci-dessous compare, pour 2023, la moyenne des chiffres d'affaires, la dette nette, l'EBITDA, le bénéfice net, et l'EBITDA/CA de 8 opérateurs majeurs de télécommunications (TELCOS) avec la moyenne de ces indicateurs pour les GAFAM. En synthèse, les GAFAM sont bien plus gros, bien moins endettés et beaucoup, beaucoup plus bénéficiaires que les TELCOS.

Tableau 2 : Comparaison de 8 opérateurs de télécommunications avec les GAFAM en 2023 (Source : Janus Henderson, rapports des sociétés).

Indicateurs ^c	TELCOS ^a	GAFAM ^b	GAFAM/TELCOS
Chiffre d'affaires	79 Mds €	293 Mds €	3,7
Dette nette	95 Mds €	42 Mds €	0,4
EBITDA	25 Mds €	85 Mds €	3,4
Bénéfice net	9 Mds €	85 Mds €	9,4
EBITDA/CA	32 %	34 %	1,1

^a TELCOS = ATT, Verizon, Comcast, Charter, Deutsch Telekom, Telefonica, Vodafone, Orange.
^b GAFAM = Alphabet, Apple, Microsoft, Amazon, Meta.
^c Taux de change : 1 US\$ = 0,91 €.

Plus précisément, les GAFAM affichent un chiffre d'affaires en moyenne 3,7 fois supérieur aux TELCOS, un EBITDA 3,4 fois supérieur, ce qui révèle un *ratio* de la rentabilité d'exploitation (EBITDA/CA) des deux catégories d'acteurs globalement identique. En revanche, le bénéfice net des GAFAM est en moyenne plus de 9 fois supérieur à celui des TELCOS s'expliquant pour partie par une dette des GAFAM plus de deux fois inférieure à celle des TELCOS.

La moyenne masque une certaine disparité des situations individuelles. La Figure 4 détaille les indicateurs économiques précédents de chacune des 8 TELCOS de l'échantillon. Il est alors intéressant de repérer deux blocs distincts : les opérateurs ancrés sur le marché des États-Unis, les opérateurs dont le marché d'origine principal se situe en Europe. Les premiers affichent des chiffres d'affaires 2 fois plus importants et sont 4 fois plus endettés que les seconds.

S'agissant du chiffre d'affaires, cette disparité rappelle une évidence : les TELCOS demeurent une activité territoriale et non « globale ». Le revenu d'un opérateur dépend de la taille et du pouvoir d'achat de la population couverte par ses réseaux. Par exemple, les réseaux de Vodafone couvrent près de 1 milliard d'individus ayant un PIB annuel moyen de 17 000 euros. Les plus de 300 millions de clients de Vodafone génèrent ainsi un revenu mensuel par client (ARPU) d'environ 10 euros, soit un chiffre d'affaires annuel de 37 milliards d'euros. ATT couvre une population d'environ 300 millions d'individus avec un PIB annuel moyen de plus de 74 000 euros. Les 150 millions de clients génèrent un ARPU mensuel d'environ 62 €.

Au total, le chiffre d'affaires d'ATT s'avère 3 fois supérieur à celui de Vodafone. La taille bien plus modeste des opérateurs européens s'explique ainsi par les caractéristiques démographiques et économiques des zones couvertes par leurs réseaux. Le cas de Deutsche Telekom, opérateur d'origine européenne est à cet égard symptomatique : son implantation aux États-Unis et en Europe lui confère les fondamentaux économiques d'un opérateur plus américain qu'europpéen, aussi bien en termes de chiffre d'affaires que d'endettement.

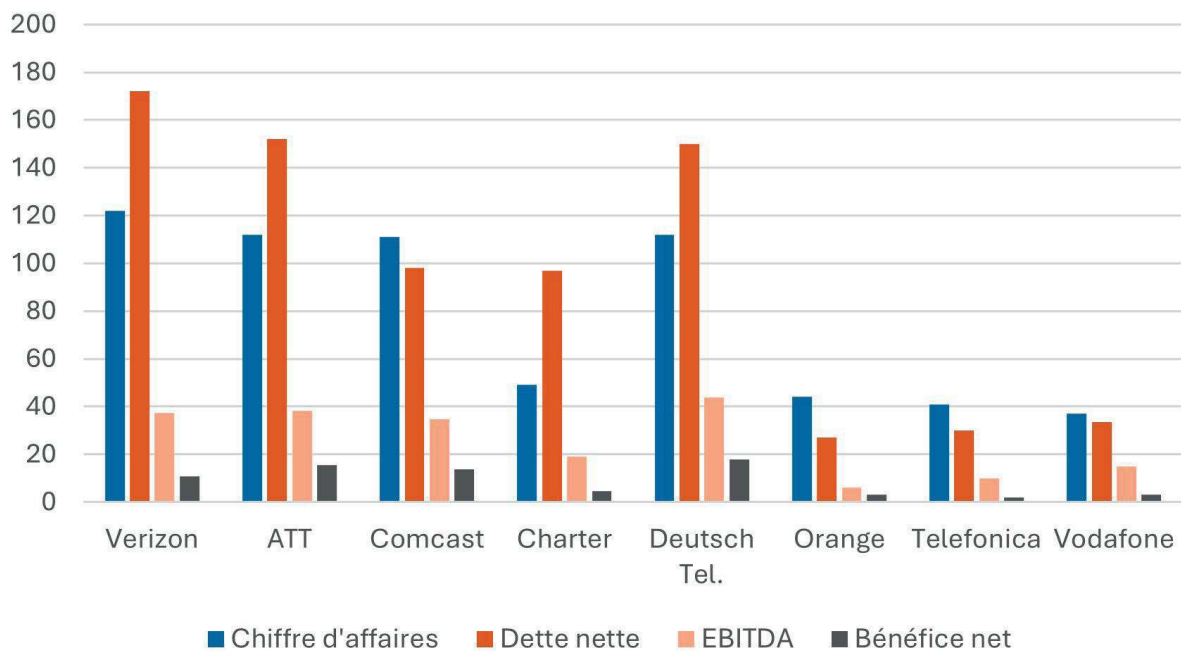


Figure 5 : Indicateurs économiques de 8 TELCOS en 2023-2024
(Source : Janus Henderson et rapports d'activité des sociétés).

Contrairement aux TELCOS, et grâce à l'interconnexion et l'interopérabilité généralisée des réseaux avec le protocole Internet, les GAFAM évoluent sur un espace de marché immédiatement mondial et non pas territorial. La très grande supériorité relative du chiffre d'affaires des GAFAM relaté ci-dessus en est la manifestation (*cf.* Tableau 2). En conséquence, à base géographique comparable, le revenu par utilisateur d'un FCA est bien moindre que celui retiré par une TELCO de ses abonnés. Par exemple, META capte un revenu moyen de 5-6 euros par utilisateur en Europe à comparer à un ARPU de 19 euros par abonné mobile ou de 38 euros par abonné fixe haut débit d'Orange en France.

Mais la contrepartie de l'ARPU d'une TELCO se trouve aujourd'hui dans le très gros investissement par client afin d'assurer un accès haut-très haut débit de qualité, un investissement sans commune mesure avec celui d'un FCA. Par exemple, selon les données de l'ARCEP, les TELCOS ont investi en France en moyenne près de 200 €/an par client fixe et mobile au cours des dix dernières années. Par comparaison, l'investissement annuel d'Alphabet par internaute français peut être évalué sur la même période à moins de 4 euros.

Ces quelques éléments d'analyse fournissent des pistes pour expliquer des divergences de fond entre le modèle économique des TELCOS et celui GAFAM observé dans le Tableau 2 *supra*.

CONVERGENCE : ESQUISSE DE PERSPECTIVES

Dans le mouvement d'intégration que représente la convergence active, les GAFAM n'ont aujourd'hui aucune incitation pour intégrer la boucle locale des réseaux de communication et devenir des TELCOS à part entière : cela pour au moins deux raisons :

- La connectivité est un service bien moins différenciable que les services de contenus et d'application. La concurrence entretenue entre TELCOS par la régulation, parfois sur des fondements plus idéologiques qu'économiques, se traduit à court-moyen terme par des taux de retour sur capitaux investis bien moindres dans l'activité des TELCOS que dans celle des GAFAM.
- Les règles de la neutralité de l'Internet garantissent un accès indiscriminé de tous les flux internet arrivant sur les réseaux des TELCOS leur interdisant *de jure* de tirer avantage du monopole d'accès aux clients que confère la boucle locale ("bottleneck monopoly"). Or le modèle économique des GAFAM se fonde sur une logique inverse de concurrence en silo avec constitution de "Walled garden", c'est-à-dire la création d'écosystèmes fermés au maximum pour contrôler l'ensemble de la chaîne de services. Cette « fermeture » concentre désormais l'attention de toutes les autorités de concurrence à travers le monde qui y voient un obstacle à l'innovation et à la concurrence.

L'Open Ran, le *edge computing*, etc., les mutations techniques en cours tendant à la virtualisation des réseaux offrent l'opportunité pour ces acteurs de s'approcher toujours plus au plus près de l'utilisateur final pour parachever leur intégration dans l'infrastructure numérique tout en laissant aux TELCOS le minimum de valeur associée aux services en les cantonnant dans un rôle d'« *utilities* », terme à comprendre dans le rôle de « service public » servant à jouer les utilités.

Les TELCOS ont jusqu'à présent échoué à s'extraire de leur métier de base. Les tentatives d'entrée dans les services *via* des protocoles particuliers (WAP ou I-mode par exemple), les diversifications dans les offres de contenu et d'applications ont été globalement des échecs (*cf.* la fusion ATT-Time Warner par exemple ou plus proche de nous les lancements peu probants de services par Orange comme Voilà, Alapage, Mappy, OCS, Orange Bank, etc.).

Le lien direct des TELCOS avec les utilisateurs finaux des services reste leur principal atout. Portant si la virtualisation des réseaux est porteuse de risques vis-à-vis des FCA, elle est tout autant porteuse de nouvelles opportunités pour les TELCOS. Le *slicing* 5G associant le SDN (réseau défini par logiciel) et la NFV (virtualisation des fonctions réseau) permettra à terme des offres différenciées de connectivité sur les réseaux public ou hybride qui seront beaucoup plus flexibles et bien moins coûteuses que les solutions actuelles.

La société Ericsson évoque ainsi la possibilité de distinguer au sein d'un même réseau 400 cas d'usage distincts relevant de 10 segments industriels différents. Toutes les sources assurent que le surcroît de revenu pour les TELCOS se chiffreraient dans les 5-10 ans à plusieurs centaines de milliards de dollars américains au niveau mondial.

Derrière cette prolifération, encore très potentielle, de connectivités différenciées se profile la perspective d'offres de services managés permettant aux TELCOS de plus grandes marges de manœuvre dans la négociation avec les tiers pour l'accès à leurs réseaux. Il reste que les autorités de régulation vont veiller à ce que l'Internet dit « ouvert » reste une composante essentielle de la bande passante disponible dans les réseaux. Car, paradoxalement, les acteurs disposant des plus grosses ressources pour négocier au mieux un accès aux services managés des TELCOS sont... les grands FCA.

De surcroît, la complexité croissante dans la gestion optimisée des différents flux transitant sur les infrastructures offrant des connectivités aux caractéristiques fortement différenciées va impliquer un recours impératif aux outils de l'intelligence artificielle : domaine où les GAFAM ne sont pas en reste.

Chasser le GAFAM par la porte, il peut revenir par la fenêtre...

De l'eau dans les nuages

Par Sylvain BOUVERET

Laboratoire d'Informatique de Grenoble
et Grenoble INP - Université Grenoble-Alpes

Aurélien BUGEAU

Université de Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800

Anne-Cécile ORGERIE

CNRS - IRISA Rennes

Et Sophie QUINTON

Centre de recherche Inria de l'Université Grenoble Alpes

Dans cet article, nous abordons la question des impacts environnementaux des infrastructures numériques sous le prisme particulier des enjeux liés à l'eau dans les centres de données (DCs) qui constituent les *clouds*. Nous tâchons de dresser un panorama de l'utilisation de l'eau dans les DCs, que cette utilisation soit directe, pour le refroidissement et l'humidification des équipements électroniques, ou indirecte pour la production électrique, la production et la fin de vie des équipements.

Ce panorama est illustré d'éléments chiffrés essentiellement issus des fournisseurs de services *cloud* eux-mêmes, afin de donner quelques ordres de grandeur sur la consommation d'eau et les tendances. L'objectif est également de discuter de ces chiffres, de la pertinence des indicateurs habituels et des transferts d'impacts potentiels associés, et de rappeler les forts enjeux spatio-temporels et de conflits d'usage liés à la ressource en eau.

INTRODUCTION

L'augmentation des niveaux de stress hydrique dans de nombreux pays représente un enjeu géopolitique majeur et un sujet de préoccupation croissante (Unesco, 2024). C'est dans ce contexte qu'émergent un certain nombre de controverses, voire de conflits, autour de l'utilisation d'eau par le secteur du numérique.

Le présent article se focalise sur les impacts sur la ressource en eau des centres de données (*data centers*, ou DCs) des infrastructures de *cloud* et présente un panorama des connaissances sur le sujet, en reprenant la structuration en *scopes* utilisée dans les bilans carbone. Ainsi, nous détaillerons d'abord l'utilisation d'eau directement dans les centres de données afin d'assurer leur bon fonctionnement (refroidissement et humidification), ce qui correspond au *scope 1*. Nous évoquerons ensuite l'eau indirectement nécessaire aux DCs pour la production de l'électricité consommée par ces DCs (*scope 2*). Enfin, nous décrirons les impacts sur la ressource en eau de la production et la fin de vie des équipements informatiques, qui relèvent du *scope 3*. Nous concluons par une discussion sur les chiffres avancés, leur pertinence et leur utilité.

LES ENJEUX LIÉS À L'EAU

Commençons par préciser quelques notions en nous appuyant sur (Maesele *et al.*, 2021). L'eau prélevée correspond à l'eau qui est extraite du milieu naturel. Une partie de cette eau est rejetée (c'est-à-dire restituée au milieu naturel) après utilisation et éventuellement traitement. L'eau consommée correspond à la différence entre l'eau prélevée et l'eau rejetée. Il s'agit d'eau évaporée, absorbée par les plantes ou le sol, intégrée dans les produits, ou transférée vers un autre bassin versant.

Les impacts sur la ressource en eau concernent d'une part la quantité d'eau consommée au regard de l'eau disponible, en prenant en compte les conditions géographiques et de saisonnalité, d'autre part les pollutions. Ajoutons que le lien entre pollution de l'eau et réduction de la disponibilité en eau ne fait pas consensus et est rarement pris en compte dans les analyses d'impact.

Depuis 2016, la norme empreinte eau ISO 14046, basée sur une approche de type cycle de vie, fournit un cadre pour rendre compte de ces impacts. Après avoir été inventoriés, tous les flux d'eau consommée et, si une approche multicritère est appliquée, de polluants émis, sont convertis en impacts potentiels. Si l'empreinte eau est monocritère, seuls les impacts dus à la privation d'eau sont évalués en utilisant des indicateurs locaux de stress hydrique.

UTILISATION DIRECTE (SCOPE 1)

Pourquoi a-t-on besoin d'eau dans un centre de données ?

L'eau utilisée dans un DC sert principalement au refroidissement et à l'humidification des équipements informatiques du centre de données. En effet, les équipements informatiques consomment de l'électricité, qu'ils dissipent sous forme de chaleur. Or, une chaleur excessive peut entraver le fonctionnement et les performances des composants. L'ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) recommande ainsi de maintenir une température à l'intérieur du centre de données comprise entre 18 et 27°C (ASHRAE TC 9.9¹). De plus, afin d'éviter les décharges électrostatiques qui peuvent survenir dans un environnement trop sec et endommager les serveurs, l'ASHRAE fournit également des recommandations concernant les plages de taux d'humidité acceptables.

Les quantités d'eau utilisées varient fortement d'un centre de données à l'autre (voir la Figure 1 page suivante), en fonction du système de refroidissement utilisé, de la localisation ou de la taille du centre de données.

Comment refroidir un centre de données ?

De nombreuses méthodes de refroidissement des DCs existent, mais le principe général reste le même (Li *et al.*, 2023) : la chaleur générée par les équipements est évacuée vers le milieu extérieur par l'intermédiaire d'un fluide, *via* l'utilisation de tours aéroréfrigérantes ou en utilisant directement l'air extérieur.

Dans les DCs refroidis par tours aéroréfrigérantes, l'eau circule en général dans deux circuits séparés, l'un fermé et l'autre ouvert. Un échangeur transfère la chaleur de l'air du DC à l'eau du circuit fermé. Cette eau chaude est ensuite envoyée au refroidisseur, qui en absorbe les calories, en général en utilisant de l'eau en circuit ouvert. Cette eau est à son tour refroidie dans des tours aéroréfrigérantes, essentiellement par évaporation.

¹ <https://www.ashrae.org/technical-resources/bookstore/datacom-series>

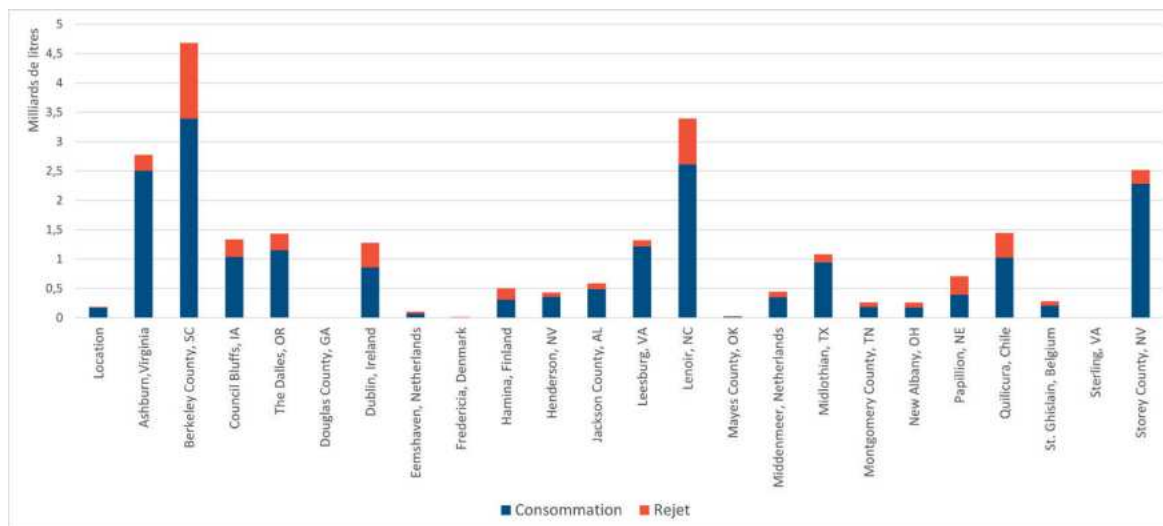


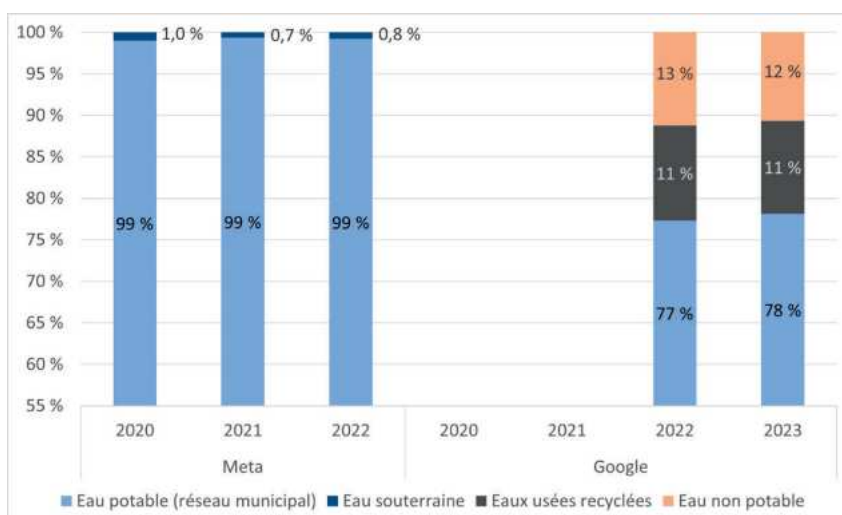
Figure 1 : Rejets et consommations d’eau dans les centres de données de Google, en 2023, en milliards de litres (Source : Google, 2023).

Dans ce système, l’eau du circuit fermé n’est pas évaporée, absorbée ou rejetée. En revanche, une partie de l’eau du système ouvert est consommée par évaporation directe. De plus, l’eau non évaporée ne peut être réutilisée qu’un nombre limité de fois (entre 3 et 10), car sa qualité se dégrade considérablement au fil des cycles. Elle est alors rejetée et de la nouvelle eau est prélevée.

Dans les DCs refroidis directement par air extérieur, le principe est encore plus simple. Lorsque les conditions le permettent, l’air extérieur est soufflé directement sur les équipements afin de les refroidir. Si l’air est trop sec et trop chaud en revanche, il peut être nécessaire d’utiliser de l’eau. Dans certaines conditions, une grande quantité d’eau peut alors être consommée par évaporation (70 % de l’eau injectée, selon le rapport de Meta, 2023).

D’autres systèmes existent pour le refroidissement des DCs, comme par exemple le refroidissement géothermique, utilisant la température constante du sous-sol ou d’une nappe phréatique pour évacuer la chaleur. La consommation d’eau de ces systèmes est quasiment nulle, mais ces méthodes de refroidissement sont peu répandues.

En résumé, l’eau des systèmes de refroidissement peut être impactée de deux manières : perdue par évaporation (dans les tours aéroréfrigérantes par exemple) ou dégradée dans son état minéral et bactériologique (ce qui nécessite un traitement). Par ailleurs, une grande partie de l’eau prélevée dans ces systèmes provient du réseau d’eau potable, même si



certain opérateurs (comme Google par exemple) utilisent aussi de l’eau issue de réseaux d’assainissement (Zhang, 2024) (cf. Figure 2 ci-contre).

Figure 2 : Source de prélèvements d’eau (Source : Meta, 2023 ; Google, 2023 ; Google, 2024).

Notons enfin que l'efficacité des systèmes de refroidissement et leur consommation d'eau est très dépendante de facteurs spatiaux et temporels. Ainsi, ces systèmes fonctionnent mieux dans des environnements où la température est moins élevée, et l'air moins sec, ce qui peut entrer en conflit direct avec des exigences de décarbonation électrique, qui typiquement exigent des lieux et des heures où l'énergie solaire est abondante (Li, 2023).

Quels sont les chiffres de consommation ?

Il est possible de connaître les quantités d'eau prélevées et consommées par certaines entreprises du numérique grâce à leurs rapports environnementaux annuels. On observe ainsi que les DCs représentent la plus grande part des consommations (et prélèvements) des géants du numérique (voir la Figure 3). Il sera par ailleurs intéressant de questionner la part de l'IA dans la croissance de l'utilisation de l'eau. En effet, des premières études (Li, 2023) estiment que l'entraînement de GPT-3 dans les centres de données de Microsoft peut consommer un total de 700 000 litres pour le *scope 1*, ce qui semble relativement faible au regard des milliards de litres d'eau consommée par les entreprises (voir la Figure 3). Par ailleurs, GPT-3 pourrait consommer 500 ml pour 10 à 50 requêtes une fois déployé, ce qui peut paraître faible, mais interroge au vu de l'utilisation croissante des IA génératives au quotidien.

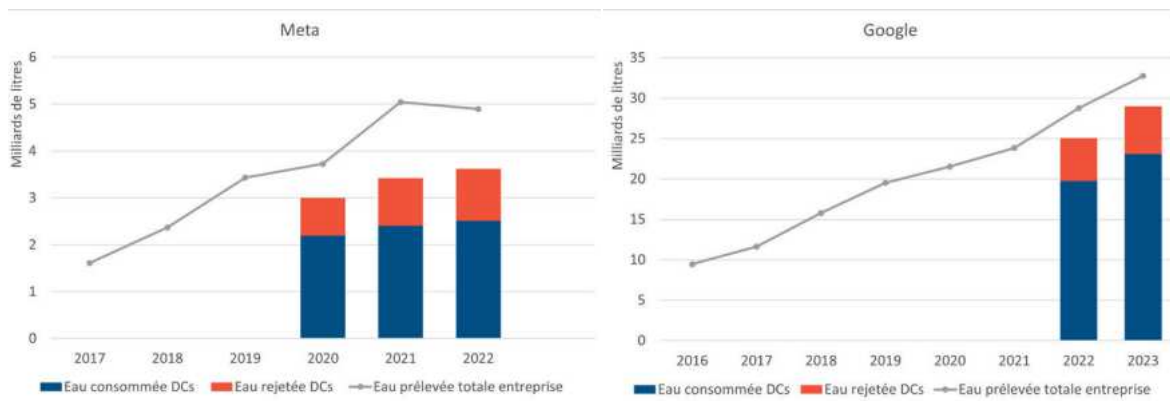


Figure 3 : Consommations et rejets d'eau en milliards de litres dans les DCs en comparaison avec les prélèvements totaux d'eau des entreprises (Source : Meta, 2023 ; Google, 2023 ; Google, 2024).

En France, un rapport ARCEP indique que le volume d'eau prélevé par les entreprises dont l'activité principale est l'exploitation de centres de données (représentant 19 opérateurs pour plus de 100 centres de données) est de 482 millions de litres d'eau en 2022 (ARCEP, 2024). C'est l'équivalent de la consommation d'eau domestique d'environ 9 000 Français.

À notre connaissance, les gestionnaires de DCs ne publient pas le résultat d'une empreinte eau exhaustive. Certaines entreprises (exemple : Meta depuis 2021) précisent si l'eau consommée provient de régions en fort stress hydrique (17 % pour Meta en 2022) ou non, et son lieu de rejet (uniquement dans le réseau d'égouts).

Un autre indicateur de performance (KPI) pour l'eau est recommandé pour les DCs : le WUE (*Water Usage Effectiveness*), spécifié dans la norme ISO/IEC 30134-9:2022 et introduit par le groupement d'entreprises Green Grid (Azevedo *et al.*, 2011). Le WUE est le ratio entre l'eau consommée par le DC et la consommation électrique des équipements informatiques. Il est exprimé en litres/kWh. Les valeurs varient fortement suivant les conditions locales : par exemple entre 0 et 9 l/kWh dans l'étude de (Karimi *et al.*, 2022) pour deux DCs en colocation dans l'agglomération de Phoenix aux États-Unis, et entre

2,04 et 2,42 l/kWh sur une étude concernant 4 DCs utilisant un refroidissement avec circuit ouvert dans quatre régions du monde dans (Gnibga *et al.*, 2024).

Pour les concepteurs et exploitants de DCs, il y a un compromis à trouver entre consommation électrique et consommation d'eau, les systèmes de refroidissement les plus efficaces en énergie reposant sur une utilisation accrue de l'eau. Les leviers d'action possibles visent notamment à adapter dynamiquement ce compromis en fonction de l'utilisation et des conditions climatiques (variations saisonnières, stress hydrique), à revoir les marges de fonctionnement en termes de redondance des infrastructures numériques, à exploiter d'autres sources que l'eau potable et enfin à réduire les usages des DCs.

UTILISATION INDIRECTE (SCOPE 2)

Au-delà de la consommation d'eau directe sur le site du DC, de l'eau est nécessaire pour la production de son électricité, là aussi souvent pour le refroidissement. Un indicateur de performance dédié (WUE_{source}) est utilisé, en prenant en compte le facteur d'intensité eau de l'énergie. Cette intensité eau, exprimée en l/kWh, varie fortement en fonction de la source d'énergie et concerne principalement de l'eau consommée par évaporation. Par exemple en France, elle est de 1,55 l/kWh pour le biogaz, 2,42 l/kWh pour le nucléaire et nulle pour le photovoltaïque (WRI, 2020). La moyenne de l'intensité eau sur la période 2016-2020 en France est de 0,87 l/kWh d'après EDF², et de 3,67 l/kWh d'après (Reig *et al.*, 2020). Le WUE_{source} représenterait ainsi en France 2,8 fois le WUE sur site (Gnibga *et al.*, 2024).

L'eau utilisée pour le refroidissement dans les *scopes* 1 (DC en lui-même) et 2 (production d'électricité) est prélevée à un endroit et un instant donné et ces prélèvements peuvent entraîner des conflits d'usage. En effet, l'eau est indispensable pour d'autres secteurs comme l'agriculture, le résidentiel, l'industrie, etc., en particulier en période de fortes chaleurs, souvent corrélées à des périodes de sécheresse. De plus, il est indiqué dans (Maesele *et al.*, 2021) que bien que l'eau consommée rejoigne le grand cycle de l'eau et donc n'entraîne pas de privation d'eau à l'échelle globale, « dans les zones arides et semi arides cette privation d'eau est effective puisque les quantités d'eau concernées ne précipitent pas sur place mais dans des zones tempérées ou sur les océans ».

LES AUTRES PHASES DU CYCLE DE VIE DES ÉQUIPEMENTS (SCOPE 3)

Le *scope* 3 concerne toutes les autres étapes du cycle de vie du centre de données, telles que la construction du bâtiment ou la fabrication et la fin de vie des équipements informatiques. Ces étapes sont largement méconnues et il n'existe à notre connaissance pas de données spécifiques aux DCs concernant leurs impacts sur la ressource en eau. Dans la suite, nous décrivons qualitativement certains enjeux majeurs liés à l'eau qui concernent le cycle de vie des équipements informatiques en général, sans qu'il soit aujourd'hui possible de les remettre en perspective.

Extraction minière

L'eau représente un enjeu majeur pour l'industrie minière, qui produit les nombreux métaux nécessaires à la fabrication des équipements numériques en général. D'après le Columbia Center on Sustainable Investment, environ 70 % des exploitations minières des six plus grandes compagnies se situent dans des pays en situation de stress hydrique

² <https://edf.publispeak.com/document-enregistrement-universel-2020/article/154/>

(Dedryver, 2020). Or l'extraction minière nécessite de grandes quantités d'eau pour les phases de broyage et de concentration du minerai. D'autre part, les déchets miniers contiennent souvent des substances dangereuses pour la santé et les écosystèmes telles que des métaux lourds (arsenic, cyanure, mercure, etc.), qui sont naturellement présentes dans la roche exploitée ou utilisées lors du traitement du minerai. Ces substances se retrouvent parfois dans l'eau des zones d'extraction minière, en particulier à cause du drainage minier acide. Elles peuvent également entraîner des pollutions de grande ampleur en cas de ruptures des digues de lacs de résidus miniers (Roche *et al.*, 2017).

Fabrication

La fabrication des équipements informatiques nécessite des grandes quantités d'eau. En particulier, la fabrication des semi-conducteurs utilise de l'eau ultra-pure pour le rinçage des disques de silicium (*wafers*) ainsi que de l'eau pour refroidir les systèmes de gravure et alimenter les laveurs de gaz. La consommation d'eau de l'un des principaux fabricants de semi-conducteurs dans le monde, TSMC, s'élevait à 150 000 m³ par jour en 2019, pour un prélèvement de 2 957 000 m³ par jour (TSMC, 2019). Cela a entraîné des problèmes majeurs pour Taïwan en 2021, année où la sécheresse exceptionnelle sur l'île a nécessité des arbitrages entre les usages (Roussilhe, 2021).

Fin de vie

Les déchets électroniques sont le flux de déchets solides qui connaît la croissance la plus rapide au monde (Baldé *et al.*, 2024). Moins d'un quart (22,3 %) de la masse de déchets électroniques est documentée comme étant correctement collectée et recyclée en 2022. Les fins de vie non régulées entraînent des risques potentiellement sévères sur la santé humaine liés à la contamination des sols, de l'air et de l'eau proches des sites de traitements ou de stockage des déchets (Ficher *et al.*, 2023) car les déchets électroniques contiennent des additifs toxiques ou des substances dangereuses.

DISCUSSION

Il paraît difficile de tirer des conclusions des chiffres incomplets et épars présentés dans cet article, suivant des indicateurs divers et souvent impossibles à remettre en perspective.

Ainsi, le WUE mis en avant par les fournisseurs d'infrastructures est un indicateur de performance relatif à la consommation énergétique, qui ne fournit pas d'information sur la consommation, les prélèvements ni les pollutions de l'eau causés par un DC. Les chiffres bruts eux-mêmes, tels que la consommation d'eau exprimée en m³ par an, ne disent rien sur les enjeux locaux tels que les lieux de prélèvement, le niveau de stress hydrique local, les conflits d'usage, etc. La prudence est donc de mise lorsqu'on observe que les quantités prélevées et consommées sont très variables d'un centre de données à l'autre, et même d'un mois à l'autre.

Parmi les améliorations envisageables pour l'analyse des impacts des DCs sur la ressource en eau, notons les propositions d'inclure la consommation d'eau verte et grise tout au long du cycle de vie et de la chaîne d'approvisionnement (Ristik *et al.*, 2015), de prendre en considération la disponibilité d'eau au niveau régional (ou national) ou une empreinte eau exhaustive. Rappelons ici que pour qu'une évaluation environnementale ait un effet positif, les informations et les connaissances qu'elle génère doivent permettre de déboucher sur des actions (Ekvall, 2019). Ainsi, l'objectif, le public cible, les scénarios de transition et le contexte dans lequel elle s'inscrit doivent être explicités et questionnés (Ekchajzer *et al.*, 2024).

Malgré la difficulté à interpréter les chiffres disponibles, les tendances de forte croissance du numérique au niveau mondial, dans un contexte d'augmentation des niveaux de

stress hydrique dans de nombreuses régions du monde, posent indéniablement problème. Cela se reflète dans l'émergence de multiples conflits locaux autour de l'installation des centres de données comme par exemple en Uruguay (Livingstone, 2023) ou en Espagne (*Courrier International*, 2023). Citons également le cas du DC Microsoft implanté sur la municipalité de Hollands-Kroon aux Pays-Bas, dont la consommation initialement estimée à 12 à 20 millions de litres d'eau par an s'était élevée dans les faits à 84 millions de litres d'eau en 2021 (DCD, 2022). Microsoft a indiqué qu'une partie importante de cette consommation était imputable à la fabrication du DC. Cependant, les oppositions locales, notamment des agriculteurs, ont conduit la province du Noordholland à contester le permis de construire octroyé par la mairie et, finalement, le gouvernement hollandais à imposer un moratoire de 9 mois sur la construction de nouveaux DCs. Il serait ainsi sans doute pertinent de mieux décrire et caractériser les enjeux locaux liés à l'eau, et leur interaction avec d'autres enjeux locaux et globaux comme l'énergie, en s'appuyant entre autres sur des travaux en sociologie (Marquet, 2019).

Cet article a présenté un panorama des différents impacts des DCs sur la ressource en eau, ainsi que les enjeux et les difficultés de la quantification de ces impacts. L'utilisation croissante des DCs soulève des craintes quant à leur lien avec des difficultés d'accès à cette ressource. Notons pour conclure que ces questions dépassent le cadre des DCs et s'appliquent au numérique en général, comme illustré par les tensions liées à la production et à la fin de vie des équipements numériques dans leur ensemble.

RÉFÉRENCES

- ARCEP (2024), « Enquête annuelle "Pour un numérique soutenable" ».
- AZEVEDO D., BELADY S. C. & POUCHET J. (2011), "Water usage effectiveness (WUE): a green grid datacenter sustainability metric", *The Green Grid*, 32.
- BALDÉ C.P. *et al.* (2024), "The global e-waste monitor", UNU, ITU & ISWA.
- COURRIER INTERNATIONAL* (2023), « Espagne. En pleine sécheresse, la consommation d'eau du futur *data center* de Meta inquiète ».
- DATA CENTER DYNAMICS (DCD) (2022), "Drought-stricken Holland discovers Microsoft data center slurped 84m liters of drinking water last year", <https://www.datacenterdynamics.com/en/news/drought-stricken-holland-discovers-microsoft-data-center-slurped-84m-liters-of-drinking-water-last-year/>
- DEDRYVER L. (2020), « La consommation de métaux du numérique : un secteur loin d'être dématérialisé ».
- EKCHAJZER D., COMBAZ J., LETONDAL C., BORNES L. & VINGERHOEDS R. (2024), "Decision-making under environmental complexity: shifting from avoided impacts of ICT solutions to systems thinking approaches", *ICT4S*.
- EKVALL T. (2019), "Attributional and consequential life cycle assessment", *In Sustainability Assessment at the XXIst century*, IntechOpen.
- FICHER M., BAUER T. & LIGOZAT A.-L (2023), « Les DEEE numériques en France », hal-04098638.
- GNIBGA W. E., CHIEN A. A., BLAVETTE A. & ORGERIE A.-C. (2024), "FlexCoolDC: datacenter cooling flexibility for harmonizing water, energy, carbon, and cost trade-offs", *ACM e-Energy*.
- GOOGLE (2023), "Google environmental report".
- GOOGLE (2024), "Google environmental report".

- KARIMI L., YACUEL L., DEGRAFT-JOHNSON J., ASHBY J., GREEN M., RENNER M., BERGMAN A., NORWOOD R. & HICKENBOTTOM K. L. (2022), “Water-energy tradeoffs in data centers: a case study in hot-arid climates”, *Resources, Conservation and Recycling*, 181, 106194.
- LI P., YANG J., ISLAM M. A. & REN S. (2023), “Making AI less ‘thirsty’: Uncovering and addressing the secret water footprint of AI models”, arXiv preprint arXiv:2304.03271.
- LIVINGSTONE G. (2023), “It’s pillage’: thirsty Uruguayans decry Google’s plan to exploit water supply”, *The Guardian*.
- MAESELE C., PRADINAUD C., PAYEN S. & ROUX P. (2021), « L’empreinte eau - Mémento graphique ».
- MARQUET C. (2019), « Binaire béton. Quand les infrastructures numériques aménagent la ville », thèse de doctorat.
- UNESCO (2024), “The United Nations World Water Development Report 2024: water for prosperity and peace; facts, figures and action examples”.
- META (2023), “Meta Sustainability Report”.
- REIG P., LUO T., CHRISTENSEN E. & SINISTORE J. (2020), “Guidance for calculating water use embedded in purchased electricity”, WRI, working paper
- RISTIC B., MADANI K. & MAKUCH Z. (2015), “The water footprint of data centers”, *Sustainability*, 7(8), 11260-11284.
- ROCHE C., THYGESEN K. & BAKER E. (2017), “Mine tailings storage: safety is no accident”, UN Environment, GRID-Arendal.
- ROUSSILHE G. (2021), « Eau et puces électroniques : l’avenir climatique et industriel de Taiwan ».
- STMICROELECTRONICS (2021), « Déclaration environnementale », site de Crolles, Rapport technique, ST Microelectronics.
- TSMC (2019), “TSMC Corporate Social Responsibility Report”.
- WORLD RESOURCES INSTITUTE (WRI) (2020), “Guidance for calculating water use embedded in purchased electricity”, working paper.
- ZHANG, M. (2024), “Data Center water usage: a comprehensive guide”.

Une perspective numérique européenne

Par Emmanuel DOTARO

Thales

Le futur numérique de l'Europe passe par la 6G qui, bien qu'elle soit dans la continuité de la 5G, se caractérise par l'opportunité de faire converger dans une combinaison inédite d'innovations de rupture sur l'ensemble des champs du numérique. Comprendre ces composants, les technologies, les architectures permet d'identifier des enjeux clés pour que ces transformations bénéficient à l'Europe. La convergence entre la connectivité et le calcul, les utilisations de l'intelligence artificielle, l'intégration d'architectures 3D allant jusqu'à l'espace, la cybersécurité et enfin la valorisation par la maîtrise de la consommation du numérique sous forme de services constituent autant de domaines où l'Europe est légitime et doit être concurrentielle. Valorisation économique et défense des valeurs sociétales où l'Europe joue à l'avant-garde sont possibles en articulant notamment investissements en Recherche et Innovation, fédération des acteurs et outils réglementaires.

LE FUTUR NUMÉRIQUE DE L'EUROPE : MAÎTRISER NOTRE DESTIN

Le numérique n'est plus un secteur à part entière, il est devenu le socle de l'ensemble des activités économiques et sociétales. Pourtant, nous n'en sommes qu'aux prémices d'une transformation profonde, comparable à un passage du Moyen Âge à une ère nouvelle encore inconnue. Les technologies, les architectures et les usages évoluent à un rythme soutenu, et cette tendance ne va qu'en s'accélégrant, potentiellement au risque d'une fracture au regard de nos valeurs et des enjeux communs de durabilité.

Les réseaux 5G et 6G, bien plus que de simples réseaux mobiles, ouvrent la voie à une multitude de systèmes, services et applications innovants. Ils constituent notamment l'infrastructure indispensable à l'exploitation du plein potentiel des données et de l'intelligence artificielle. La 6G, en s'appuyant sur la 5G, va encore plus loin en intégrant, dans une architecture inédite, des capacités de *cloud (edge, cloud continuum)*, de connectivité intelligente, de sécurité différenciée, d'intégration de l'espace et de convergence avec le "sensing". Le théâtre de la 6G se confond donc plus largement avec celui de notre perspective numérique.

Que ce soit au sein des différents champs technologiques, de leur intégration vers des systèmes publics et privés ou des usages sous forme de services, l'Europe est *de facto* légitime et se doit d'être acteur réel des transformations maîtrisées.

Les enjeux économiques sont colossaux. Il s'agit de plusieurs centaines de milliards d'euros d'investissements à l'échelle mondiale, et la position de chaque acteur sur l'échiquier mondial en dépend. Dans cet environnement compétitif, l'Europe, met un accent différenciateur sur un développement durable, à la fois économique, environnemental et sociétal. Ces considérations peuvent quelquefois être perçues, voire décriées (par ceux qui

ont intérêt à le faire) comme des freins à l'innovation. Elles sont souvent vues au-delà des valeurs intrinsèques, comme garantissant l'acceptation sociétale et sécurisant les investissements financiers, y compris au-delà de nos frontières.

Le point de vue européen est largement couvert dans « La décennie numérique de l'Europe »¹, adoptée par l'Union européenne qui définit des objectifs ambitieux pour 2030. Il vise à placer les citoyens et les entreprises au cœur de la transformation numérique, tout en promouvant des valeurs centrées sur l'humain et des critères de durabilité stricts.

Mais au-delà de ces objectifs, c'est la capacité de l'Europe à maîtriser son destin numérique et ses impacts qui est en jeu. Cette souveraineté ou autonomie numérique n'a de sens que si nous comprenons pleinement les enjeux et fondamentaux des transformations. Face à la puissance des acteurs d'autres régions, chinois, américains et très prochainement indiens, il est impératif que l'Europe adopte une stratégie réfléchie et ambitieuse.

La 5G est en déploiement rapide pour le marché grand public et adoptée au rythme des cycles de vie spécifiques des secteurs verticaux, et la 6G est prévue pour 2030. L'avenir numérique de l'Europe se construit dès aujourd'hui.

LES 5G ET 6G, LES FUTURS « XG » SONT DÉTERMINANTS POUR LE FUTUR NUMÉRIQUE DE L'EUROPE

En 2024, alors que la 5G se déploie à grande échelle, la question de l'avenir numérique de l'Europe se pose de façon imminente. La 6G, prévue d'ici 2030, bouleverse le paysage du secteur et de ses utilisateurs, elle ouvre de nouvelles perspectives pour le continent.

Loin d'être un simple progrès incrémental, la 6G représente une évolution conséquente de la 5G. Elle ne se limite pas aux réseaux d'accès radio, mais, au regard des initiatives de recherche et innovation majeures à travers le monde, englobe un vaste éventail de champs numériques. Ce consensus est illustré dans la Figure 1 regroupant différents usages (dual), architectures distribuées, privées et publiques (dont des réseaux non-cellulaires, le *cloud continuum*, l'Internet des Objets et 3D), et technologies sous-jacentes (dont l'intelligence artificielle).

La 6G est l'objet de programmes de recherche et innovation substantiels, c'est le cas en Europe avec l'entreprise commune "Smart Networks and Services" (SNS)² dont le programme se base sur un agenda de recherche³ (NetworldEurope, 2022) recoupant largement les champs cités ci-dessus.

Les États-Unis^{4 5} sont engagés au travers de différents programmes, organisations et associations pour des applications civiles et militaires, la Chine⁶, l'Inde⁷, le Japon⁸, la

¹ Décennie numérique de l'Europe : objectifs pour 2030, Commission européenne (europa.eu), https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_fr

² Missions & Objectives - SNS JU (europa.eu), <https://smart-networks.europa.eu/missions-and-objectives/>

³ Microsoft Word - SRIA 2022 Technical Annex 20221208 - AFPC - cleaned.docx (sns-ju.eu), <https://bscw.sns-ju.eu/pub/bscw.cgi/d95665/SRIA%202022%20Technical%20Annex%20Published.pdf>

⁴ Next G Alliance, <https://nextgalliance.org/>

⁵ FutureG Home – DCTO(S&T), <https://rt.cto.mil/futureg-home/>

⁶ Internet Access (imt2030.org.cn), <https://www.imt2030.org.cn/html/default/yingwen/Introduction/>

⁷ Bharat 6G Alliance, <https://bharat6galliance.com/bharat6G/>

⁸ XG Mobile Promotion Forum XGMF, <https://xgmf.jp/en/>

Corée, quelquefois en collaboration avec des entités européennes⁹ sont aussi très largement engagés dans la course avec leur propres programmes et investissements nationaux.

La 6G est centrale dans les transformations numériques à venir, sa richesse d'éléments et sa complexité intrinsèque ne doivent pas rebuter, car elles recèlent un potentiel pour l'Europe, potentiel qui sinon serait laissé à des géants peut-être moins convaincus à l'origine par les valeurs européennes.

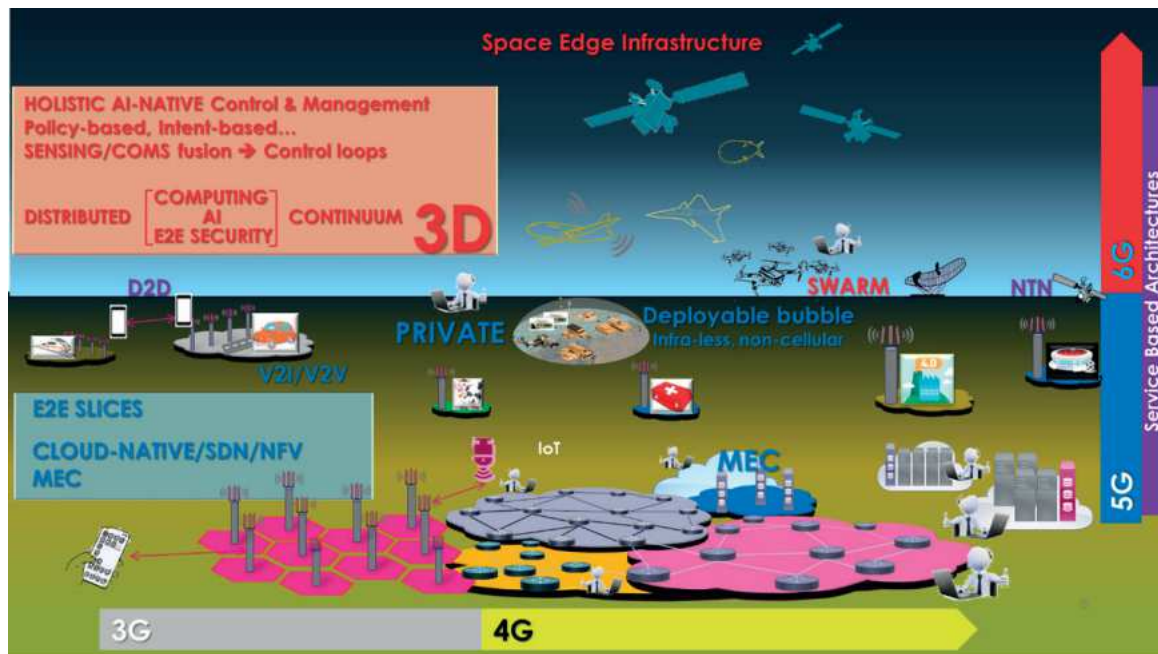


Figure 1 : Architectures xG (Source : Lancement PEPR, 2023, <https://pepr-futurenetworks.fr/>).

Au cœur de la 6G se trouve l'utilisateur, qu'il soit humain, organisation, machine ou application. Les services numériques de demain seront centrés sur ses besoins et devront explicitement exposer leurs propriétés leur permettant de faire des choix éclairés en fonction de critères tels que l'impact environnemental ou le niveau de sécurité.

Cette richesse de périmètres et technologies implique un questionnement radical des acteurs fournissant ces services dont l'origine n'est plus simplement les télécommunications au sens historique du terme.

S'il existe des variations sur l'étendue du périmètre ou les priorités, il y a bien un consensus sur les principes fondamentaux de la 6G avec notamment la prise en compte des questions de durabilité, d'ouverture ou encore d'objectif d'unicité des standards. Sur ce dernier point les efforts consentis au 3GPP¹⁰, dans de nombreux groupes relatifs aux sujets 6G (virtualisation, intelligence artificielle, sécurité, *integration sensing at commu-*

⁹ Unlocking the potential of 6G: Finland and South Korea join forces in the 6GBRIDGE-6GCORE project to develop advanced system architecture – 6G Flagship, <https://www.6gflagship.com/news/unlocking-the-potential-of-6g-finland-and-south-korea-join-forces-in-the-6gbridge-6core-project-to-develop-advanced-system-architecture/>

¹⁰ 3GPP Commits to Develop 6G Specifications, <https://www.3gpp.org/news-events/3gpp-news/partner-pr-6g>

nications...) de l'ETSI¹¹ et autres organisations de standardisation sont clés et seront aussi largement déterminants pour les retombées économiques.

Nous reviendrons sur le positionnement relatif, des forces et faiblesses de l'Europe avec une sélection de champs structurants de la 6G couvrant les technologies, les architectures et les enjeux au travers des services.

ENJEUX SPÉCIFIQUES ET IMPLICATIONS EUROPÉENNES

Edge et cloud continuum

La 5G, dite *cloud-native*, utilise une infrastructure de virtualisation pour instancier des fonctions réseaux, amorçant ainsi une consommation "as a service" des ressources réseaux. Cette évolution représente une première étape de convergence technologique, basée sur des composants logiciels prédominants et une infrastructure matérielle dépendante du *cloud*. Ce sont des domaines où l'Europe n'est pas en position de force et ne dispose pas d'acteurs à l'échelle de ses grands concurrents.

La 6G, quant à elle, promet une étape de convergence encore plus poussée entre les capacités de calcul et de connectivité. Cette vision, bien que sujette à des variations d'interprétations, anticipe un système moins centralisé que le *cloud* actuel. Les ressources de calcul et de connectivité sont distribuées et orchestrées de manière optimale pour servir les applications, en particulier celles utilisant l'intelligence artificielle fédérée.

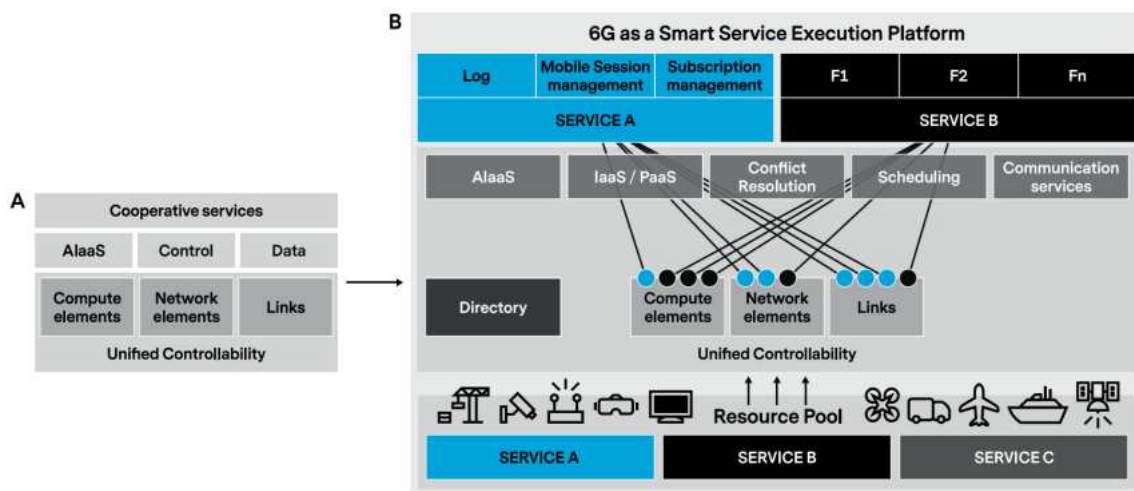


Figure 2 : Convergence réseaux et *cloud*
(Source : Livre blanc de NetworldEurope, 2022).

La Figure 2 issue du livre blanc de NetworldEurope¹² illustre cette distribution de ressources au service des applications.

L'Europe se trouve donc face à un défi majeur : comment garantir que la future infrastructure numérique ne soit pas dominée par les acteurs du *cloud* américains ou chinois ? Cette

¹¹ ETSI - Welcome to the World of Standards!, <https://www.etsi.org/>

¹² NETWORLD EUROPE, Strategic Research and Innovation Agenda 2022 (sns-ju.eu), <https://bscw.sns-ju.eu/pub/bscw.cgi/d95660/SRIA-2022-WP-Published.pdf>

question se décline immédiatement sur la souveraineté de la chaîne d’approvisionnement et de la valeur associée.

La DG CONNECT (direction générale des réseaux de communication, du contenu et des technologies) a publié en 2024 un livre blanc¹³ représentatif des besoins stratégiques de l’Europe, introduisant en réponse un concept de “3C Network” - “Connected Collaborative Computing”. Ce concept vise à créer un écosystème européen dynamique qui englobe les semi-conducteurs, les capacités de calcul en périphérie (Edge) et dans le *cloud*, les technologies radio, les infrastructures de connectivité, la gestion des données et les applications.

L’essor de l’informatique en périphérie (*edge computing*) va s’appuyer sur de nombreuses machines/terminaux disposant d’une capacité de calcul importante, y compris les robots, drones, dispositifs médicaux, appareils portables et voitures autonomes. Le traitement des données ne sera plus obligatoirement réalisé de façon hyper-centralisée et énergivore.

L’enjeu pour l’Europe est de développer une industrie compétitive capable de rivaliser sur des segments clés de la chaîne de valeur numérique et respectueuse des avancées sociétales (vie privée, durabilité) européennes.

Cette transformation, notamment celle des acteurs historiques des télécommunications, nécessite des investissements massifs, à l’échelle de ceux des grands fournisseurs de *cloud* dans les capacités *cloud*, *edge* et d’intelligence artificielle. L’Union européenne dispose de plusieurs instruments financiers pour soutenir la recherche et le développement dans le secteur des communications, comme le “Smart Networks and Services Joint Undertaking” (SNS JU) dans le cadre du programme Horizon Europe, ainsi que InvestEU, le programme pour le numérique européen (DEP) et le Mécanisme pour l’interconnexion en Europe (CEF) Digital.

L’Europe peut également s’appuyer sur des financements existants. En décembre 2023, la Commission a approuvé une aide d’État de 1,2 milliard d’euros de sept États membres pour un projet d’intérêt commun européen (IPCEI) dans le domaine des infrastructures et services de *cloud* de nouvelle génération (IPCEI CIS). De plus, un autre IPCEI a été approuvé en juin 2023 pour soutenir la recherche, l’innovation et le premier déploiement industriel de la micro-électronique et des technologies de communication (IPCEI ME/CT).

L’industrie européenne s’emploie également au développement de solutions en *open source*, notamment « Sylva »¹⁴ nourrissant le concept de Telco Cloud agissant sous la bannière de la Linux Foundation. Ces initiatives visent à réduire la dépendance aux composants logiciels et matériels absents aujourd’hui en Europe.

En mobilisant ses acteurs et ses ressources, en investissant dans la recherche et le développement et en promouvant des solutions *open source*, l’Europe peut relever les défis de la convergence technologique et créer un réseau collaboratif, ouvert et fédéré de calcul connecté compétitif et concurrentiel.

Enfin, l’Europe se dote d’outils de régulation, et de certification dans le domaine du *cloud*. En cours de finalisation, le schéma EUCS (pour European Union Cybersecurity Certification Scheme for Cloud Services) pourrait intégrer, malgré la réticence des acteurs du

¹³ White Paper - How to master Europe’s digital infrastructure needs? Shaping Europe’s digital future (europa.eu), <https://digital-strategy.ec.europa.eu/en/library/white-paper-how-master-europes-digital-infrastructure-needs>

¹⁴ White Paper Operators Telco Cloud, https://gitlab.com/sylva-projects/sylva/-/blob/main/White_Paper_Operators_Sylva.pdf

numérique américains¹⁵, la transposition de la certification SecNumCloud¹⁶ élaborée par l'ANSSI.

Les architectures 3D

La compétition mondiale est aussi particulièrement intense dans le domaine du *New Space*. L'Europe, consciente des enjeux stratégiques et économiques de ce domaine, se positionne comme un acteur majeur grâce aux actions de l'agence spatiale européenne (ESA).

L'un des défis clés du *New Space* est justement le développement de la 6G¹⁷.

L'ESA joue un rôle crucial dans le développement de la 6G en Europe en soutenant des projets de recherche et développement, en encourageant la collaboration entre les acteurs publics et privés, et en définissant une vision stratégique pour le futur des communications et systèmes intégrant un segment spatial. Au-delà des technologies de communication radio ("Non Terrestrial Networks", NTN) ou encore d'optique en espace libre. Il s'agit bien d'intégrer et transposer les concepts de la 6G, notamment par le biais d'un "Space Edge Computing" ou encore de déploiement de capacités d'intelligence artificielle :

- le programme ARTES qui vise à développer des technologies de communication spatiales de nouvelle génération, y compris des systèmes 6G ;
- le programme Iris qui vise à établir une constellation de satellites en orbite basse pour fournir des services de connectivité haut débit à l'échelle mondiale ;
- le programme Catalyst qui soutient des projets innovants utilisant des technologies spatiales, y compris des applications 6G.

L'IA pour la 6G, la 6G pour l'IA

L'intelligence artificielle (IA) et la 6G sont deux ensembles de technologies qui promettent de révolutionner de nombreux secteurs. L'IA, avec ses capacités d'apprentissage et d'analyse de données, de contrôle/gestion autonome peut optimiser les performances de la 6G, tandis que la 6G, avec son réseau ultra-rapide et performant, l'orchestration intelligente et concomitante des ressources de connectivité, calcul, sécurité peut favoriser l'essor de l'IA distribuée.

La 6G est dite « AI-native », elle comprend de nombreux champs d'application de l'IA :

- communications sémantiques : du capteur à la transmission frugale de l'information ;
- couches radio : optimisation des paramètres réseau, allocation dynamique des ressources spectrales, prédiction et correction des interférences ;
- réseau cœur : gestion de la congestion, routage intelligent, orchestration des fonctions et ressources réseau, calcul, sécurité, détection et résolution d'anomalies. Maintenance prédictive... ;
- interface avec les utilisateurs : suivant des concepts "d'intent-based" pouvant recourir à de l'extraction sémantique en langage naturel ;

¹⁵ 05.25.2023 Industry Letter on EUCS, <https://www.uschamber.com/assets/documents/Joint-U.S.-Industry-Letter-on-EUCS.pdf>

¹⁶ SecNumCloud pour les fournisseurs de services *cloud*, ANSSI, <https://cyber.gouv.fr/secnumcloud-pour-les-fournisseurs-de-services-cloud>

¹⁷ ESA, 6G and Satellites: Intelligent Connectivity for a Sustainable Future, https://connectivity.esa.int/sites/default/files/ESA_6G_White%20Paper_Dev_Proof_V14..pdf

- services : personnalisation des services, expériences utilisateur immersives, réalité augmentée et virtuelle optimisées, cybersécurité renforcée.

La difficulté pour l'Europe est de diriger les efforts consacrés à l'IA vers des applications inspirées par les besoins de la 6G. Les cycles de mode de l'IA ne sont pas aujourd'hui motivés par les acteurs traditionnels de la 6G. La conséquence directe est encore une sous-exploitation des avancées propres à l'IA pour des mises en œuvre dans le domaine de la 6G.

La 6G permet le déploiement de l'IA, en particulier dans des environnements distribués au plus près des terminaux pour une meilleure réactivité et une réduction de la charge du réseau central. Là encore, il y a un déficit de transversalité et une nécessité de compréhension partagée des défis pour que les différentes formes d'IA et la 6G s'alimentent mutuellement... L'Europe, en investissant et en coordonnant ces domaines clés, peut se positionner comme un *leader* dans la transformation numérique globale.

Génération concomitante de cybersécurité

L'avènement de la 6G s'accompagne de défis majeurs en matière de cybersécurité, qu'il est indispensable de relever pour garantir la protection des données, la résilience des infrastructures critiques pour la sécurité globale des citoyens et du tissu économique. La confiance ne peut se donner *a priori* et n'est que la résultante des capacités qui seront construites pour la protection, détection, réponse et gouvernance de la sécurité.

La responsabilité de l'Europe est de transposer l'excellence de ses avancées^{18 19} en cybersécurité, en les positionnant face aux besoins de la 6G et anticipant les besoins d'une valeur ajoutée dans ce domaine pour les utilisateurs.

L'Europe doit en premier lieu être coopérative et montrer par sa gouvernance une coopération active sur le partage des vulnérabilités et plus généralement de la "Cyber Threat Intelligence", de la coordination de la réponse à incident.

L'Europe doit garantir que ses valeurs sont respectées et que la 6G n'introduit pas de brèches, notamment pour les données privées mais plus largement pour les données sensibles liées à son économie et sa sécurité.

L'Europe doit se positionner comme champion d'une nouvelle génération de cybersécurité, actant la fin des défenses périmétriques et statiques pour des systèmes complexes et parfaitement dynamiques !

Cela implique la prise en compte défis spécifiques de la cybersécurité de la 6G.

Évaluation des niveaux de sécurité

La complexité croissante des réseaux 6G rend l'évaluation des niveaux de sécurité un défi majeur. De nouvelles méthodes et outils d'évaluation sont nécessaires pour identifier et quantifier les risques potentiels. Ces méthodes devraient notamment permettre d'exposer les niveaux de sécurité respectifs des différents services numériques et d'éclairer les utilisateurs sur leurs risques.

¹⁸ Règlement général sur la protection des données (RGPD), <https://eur-lex.europa.eu/FR/legal-content/summary/general-data-protection-regulation-gdpr.html>

¹⁹ Cybersécurité des réseaux et des systèmes d'information (2022), <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=LEGISSUM:4637829> ; The EU Cybersecurity Act, <https://eur-lex.europa.eu/EN/legal-content/summary/the-eu-cybersecurity-act.html> ; Cyber Resilience Act, Shaping Europe's digital future, <https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act>

Aspects opérationnels dynamiques

La 6G introduira de nouveaux défis opérationnels en matière de cybersécurité, notamment la gestion d'un grand nombre d'appareils connectés, la détection des menaces en temps réel et la réponse aux incidents.

Architectures Zero Trust et Confidential Computing

L'adoption d'architectures Zero Trust et de Confidential Computing (plus généralement des concepts de "Data Centric Security") est essentielle pour renforcer la sécurité de la 6G. Ces approches reposent aujourd'hui sur une offre trop peu européenne.

Sécurité "as a service"

La mise à l'échelle des déploiements de la cybersécurité impose un maintien à l'état de l'art inaccessible et économiquement intraitable pour les profanes ; sont concernées sur le seul périmètre européen plus de 20 millions d'entreprises. La 6G, avec son architecture orientée service, doit bénéficier en Europe du développement d'une offre à intégrer, que ce soit pour des solutions publiques ou privées.

Les services : au risque de perdre ses valeurs

Notre futur numérique se dessine sous la forme de services, consommés à la demande, accessibles à tous ("Everything as a Service"). Cette évolution place l'interface utilisateur (humain, machine ou logiciel) au cœur de l'expérience, et les acteurs qui maîtrisent ces interfaces seront les mieux placés pour réussir.

Or, aujourd'hui, ce maillon stratégique est dominé par de grands acteurs hors Europe. Face à ce constat, il est urgent que l'Europe reprenne l'initiative sous peine de perdre toute possibilité de retour.

La maîtrise de l'interface est essentielle pour contrôler les chaînes de valeur. C'est elle qui permet de définir comment les services sont présentés aux utilisateurs, en mettant en avant leurs attributs et leurs différenciations, qu'ils soient :

- économiques (distribution de la valeur dans les diverses chaînes de fournisseurs) ;
- sociétaux (respect de la vie privée, de l'éthique, de la propriété intellectuelle) ;
- ou environnementaux (impacts énergétiques, cycle de vie des systèmes).

Or, de nombreux services numériques actuels sont des « boîtes noires », opaques et peu transparentes. Il est difficile, voire impossible de distinguer le positionnement réel dans le système de valeurs. Dans ce contexte, comment les entreprises européennes, même les plus performantes, peuvent-elles espérer un retour sur investissement si elles sont soumises à des plateformes tierces aux intérêts divergents ?

La constitution de l'offre de services numériques visible des utilisateurs repose sur l'agrégation et la composition complexe d'un très grand nombre de composantes, convergées, allant des réseaux d'accès au *cloud*, services de sécurité, l'intelligence distribuée et une infinité de fonctionnalités numériques potentielles. Maîtriser la distribution des valeurs impose *de facto* la maîtrise des plateformes d'accès aux services ainsi que les politiques d'intégration des fournisseurs.

La Figure 3 illustre l'intégration de défis permettant la valorisation des performances des constituants des futurs systèmes numériques quelle que soit leur couche.

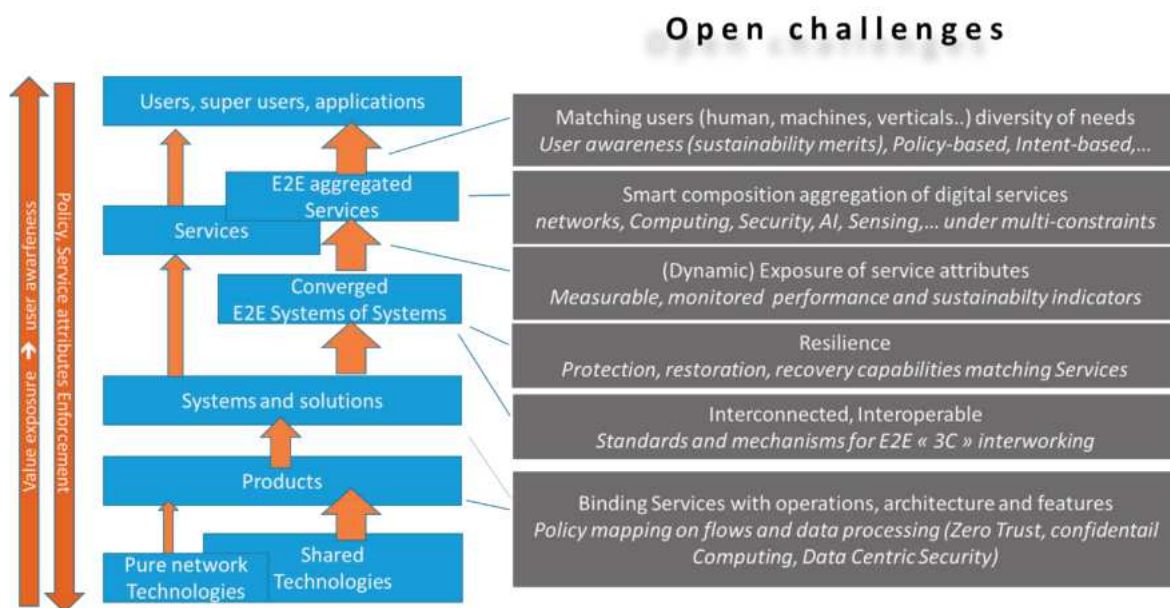


Figure 3 : Capture de la valeur : défis relatifs aux services
(Source : Préparation du programme de travail SNS JU, 2024).

C'est pourquoi l'Europe devrait se doter d'une plateforme ouverte et régulée d'accès aux services numériques. Cette plateforme, conforme aux valeurs et au droit européens, permettrait de :

- Différencier les niveaux de service en fonction des standards, réglementations et certifications en vigueur en Europe. Cela permettrait notamment de mettre en avant les services les plus respectueux de l'environnement, des exigences de niveaux de sécurité et de la vie privée.
- Stimuler l'innovation et la création de nouveaux services en garantissant un accès ouvert à la plateforme pour des acteurs sélectionnés sur la base de valeurs européennes. Cela permettrait notamment de réduire les barrières à l'entrée pour les petites entreprises et de garantir la présence d'acteurs dans les domaines critiques.

La création d'une telle plateforme est une opportunité pour l'Europe de façonner un avenir numérique pour ses citoyens et ses entreprises. La construction d'une trajectoire crédible pourrait s'appuyer sur les efforts de Recherche et d'Innovation dirigés au niveau national et européen, cela permettrait notamment la capitalisation et la valorisation des avancées dans de très nombreux domaines techniques et scientifiques sous forme de services. Elle devrait s'inscrire dans une stratégie d'application sectorielle répondant aux différents besoins du marché, notamment ceux relevant d'exigences applicables aux infrastructures critiques.

QUELLE AMBITION, QUELLE STRATÉGIE ?

L'avenir numérique de l'Europe, intégrant le rôle crucial des technologies et architectures 6G dans la transformation de tous les secteurs de la société offre un panorama ambitieux et positif. Pour saisir pleinement les opportunités offertes par la 6G, l'Europe doit impérativement relever un ensemble de défis majeurs :

- Concurrence technologique : l'Europe doit se positionner comme un acteur de premier plan dans la recherche et le développement de la technologie 6G, en rivalisant avec les puissances technologiques mondiales telles que les États-Unis et la Chine.

- Souveraineté numérique : l'Europe doit garantir sa souveraineté numérique en développant ses propres infrastructures, outils et services numériques, en s'appuyant sur ses valeurs et sa régulation.
- Valorisation des services : l'Europe doit aller au-delà de la simple production de technologies pour se concentrer sur la création de valeur pour ses citoyens et ses entreprises en développant des services innovants et utiles tout en maîtrisant l'accès à ces services.
- Cybersécurité : l'Europe doit garantir la sécurité et la fiabilité de ses infrastructures numériques face aux menaces croissantes en matière de cybersécurité.
- Investissements massifs dans la recherche et le développement : l'Europe doit poursuivre et accroître ses investissements dans la recherche et le développement de la technologie 6G. La capitalisation de son excellence technique doit se faire par des approches fédératives pour une mise à l'échelle de la compétition globale.
- Développer des champions européens : l'Europe doit soutenir l'émergence de champions européens dans le domaine des services numériques, capables d'offrir les environnements d'intégration et composition de services pour tout type de fournisseur.

La perspective numérique de l'Europe n'est pas sombre si elle adopte une stratégie ambitieuse, capitalisant intelligemment sur son excellence et ses valeurs.

La 6G : promesses et défis à l'horizon 2030

Par David GESBERT

Directeur d'EURECOM et professeur en Systèmes de Communications

Cet article fait un point rapide sur certaines attendues de la 6G semblant faire consensus à ce jour. Nous donnons aussi un éclairage sur les concepts scientifiques qui joueront un rôle important dans ces futurs réseaux sans fil, tels que les surfaces réfléchissantes, l'intelligence artificielle, la spatialisation de l'information et le *sensing* intégré à la communication. Nous relevons enfin certains défis sociétaux qui conditionneront le succès de la 6G tels que la sécurité informatique, le droit à la vie numérique privée, l'impact écologique de ces systèmes.

INTRODUCTION

L'arrivée de la 6G, prévue à l'horizon 2030, sera marquée par une nouvelle norme sous l'égide du 3GPP, avec plusieurs étapes de progression sous la forme de mise à jour du standard (*releases* dans la terminologie 3GPP).

Que peut-on déjà dire de la 6G en 2024 ? En réalité, le débat est encore trop vif pour annoncer avec certitude les cas d'usage phares de la 6G ou les briques de *design* qui seront choisies pour lui servir de support, puisque la technologie est encore en phase de R&D. L'objectif de cet article est donc double. D'abord faire un point rapide sur certaines attendues semblant faire consensus à ce jour. Deuxièmement, donner un éclairage sur les concepts scientifiques qui joueront un rôle important dans ces futurs réseaux sans fil, en tenant compte aussi de certains défis sociétaux qui conditionneront le succès de la 6G.

PROMESSES ET ATTENTES DE LA 6G

Les nouveaux modes d'utilisation : réalités virtuelles et robots connectés

En incluant les modes d'utilisation industriels et ceux liés à l'internet des objets (IoT) en plus de son rôle classique de tuyau d'accès à l'internet mobile haut débit, la 5G a été la première norme de communication sans fil à devoir prendre en considération l'équilibre des ressources spectrales pour servir en simultanément les besoins distincts d'utilisateurs humains et non humains, notamment grâce aux techniques de *slicing*. La 6G va faire face aux mêmes défis avec cependant des besoins rehaussés de qualité de service attendue.

Du côté des utilisateurs humains, les modes d'utilisation évoluent vers la communication dite immersive par laquelle les données de contexte (environnement vidéo 3D réel ou synthétique) sont transportées dans le canal radio aux côtés des données usuelles (voix) telles que dans la communication holographique. Par ailleurs l'utilisation accrue du *metaverse* a des fins éducatives ou ludiques par le biais des casques de réalité virtuelle ou augmentée aux performances toujours plus impressionnantes crée aussi de nouveaux cas d'usage très haut débit pour la communication sans fil [1].

Côté machines, avec la généralisation des robots connectés, véhicules autonomes guidés par le réseau, capteurs connectés de toute nature, les usages non directement sous le contrôle d'un humain se multiplient. Ainsi la 6G verra la transition vers les réseaux *machine-centric* s'accélérer. Aux capteurs connectés bas débit déjà servis par la 4G et la 5G (capteurs trafic routier, pollution, météo, agriculture, certains appareils électroménagers du quotidien, etc.), viendront prochainement s'ajouter des clients machines beaucoup plus exigeants sur la qualité du lien radio, notamment dans le contexte de la robotique. On peut par exemple citer le cas des drones sans pilote utilisant le réseau sans fil pour les algorithmes de guidage, la voiture autonome utilisant la fusion de données locales issues des capteurs embarqués (radar, lidar) avec les informations plus globales provenant du réseau cellulaire, et enfin les robots (ou « cobots » pour *collaborative robots*) industriels communiquant entre eux pour la résolution de tâches mécaniques complexes dans le contexte de l'usine du futur. Tous ces cas d'utilisation machine posent des défis considérables pour les futurs réseaux de communication sans fil [2].

Réseaux hautes performances

Pour rendre les modes d'utilisation décrits ci-dessous réalisables, la future norme de communication sans fil devra relever le niveau de qualité afin de dépasser les performances des réseaux 5G actuels [3]. Le besoin de transporter d'importants flux vidéo générés par la réalité virtuelle ou augmentée, ou par des capteurs de vision très haut débit sur les robots (drones ou voitures) conduit à relever la qualité du lien radio de manière significative. La qualité est ici à la fois mesurée par le débit maximum atteignable pour lequel l'objectif de 1 TB/sec est donné (dans les conditions où toute la ressource spectrale est assignée à un utilisateur seul) mais aussi par le débit moyen mesuré en conditions réelles, pour lequel aucun chiffre ne peut être avancé à ce stade puisqu'il dépend de la densité d'utilisation de déploiement du réseau. La capacité, elle, est mesurée par le nombre de liens avec une performance cible qu'il est possible de maintenir en simultané par m^2 ou km^2 . En intégrant les modes d'utilisation issus de l'internet des objets aux côtés des terminaux humains classiques (téléphones, tablettes), le chiffre moyen de 1 utilisateur radio par m^2 peut être atteint. Enfin, la connectivité globale sur un maximum de surface de la planète est aussi un enjeu de choix nécessitant l'évolution de la norme, considérant notamment les technologies dites non terrestres (NTN) tels que les communications satellites et les plateformes de haute altitude pour les intégrer au réseau cellulaire.

Pour les cas d'usages liés à la cobotique (collaboration homme-robot et robot-robot), aux critères classiques de débit s'ajoutent des métriques supplémentaires telles que la latence et le positionnement, déjà présente pour la 5G mais destinées à être considérablement renforcées.

La latence mesure la réactivité du réseau par le temps écoulé entre une commande lancée au transmetteur et le moment où l'action peut commencer à être exécutée au récepteur. Mis en avant dans le cadre de la 5G et des réseaux dits *time-sensitive networks*, mais tardant encore à se matérialiser par des produits disponibles sur le marché, des protocoles de communication dédiés pouvant délivrer une latence de l'ordre de la msec ou même inférieure devront être mis en œuvre par la 6G. Une latence très faible étant particulièrement nécessaire au bon fonctionnement des cobots, ou pour les cas d'usage les plus sensibles de l'internet tactile, ou des capteurs haptiques sont connectés à des nœuds radio.

En plus de la latence, la robotique connectée pose des défis de géolocalisation qui ne peuvent être traités par les seuls systèmes satellitaires (GNSS dont le fameux GPS fait partie) à cause de leur manque de précision et leur incapacité à traiter les situations où la propagation satellitaire est obstruée (par exemple : environnement urbain dense ou *indoor*). Dans le contexte de l'usine du futur, il est essentiel pour des robots collaboratifs industriels enchaînant des tâches mécaniques de précision de pouvoir s'auto-positionner pour assurer la sécurité des opérations. On vise alors une précision de l'ordre du cm. On

retrouve des exigences similaires pour les cas où humains et robots interagissent directement, dans la même zone, par exemple dans une usine. Pour ces raisons, la géolocalisation de très haute précision fait partie des critères de performances faisant l'objet d'une attention particulière et qui va requérir de nouvelles méthodologies pour la prochaine génération.

LES BRIQUES TECHNOLOGIQUES DE LA 6G

Plusieurs pistes technologiques sont à l'étude pour répondre aux attentes de la 6G en termes de performances. Nous brossons ici un tableau rapide, non exhaustif, des principaux concepts.

Les fréquences

LA 6G va tout d'abord se reposer sur les fondements de la 5G, et cela inclus l'utilisation combinée de plusieurs plages fréquentielles. La zone sous les 6 GHz restera clé pour sa capacité à faciliter la propagation en environnement urbain et pénétrer les structures pour la couverture *indoor*. Le déploiement des ondes millimétriques (30 à 300 GHz) qui s'est révélé très difficile en 5G car très sensible aux obstacles se poursuivra pour des cas d'utilisation limités, tout comme l'introduction du THz (bandes de 300 GHz à 3 THz) qui peut donner d'excellentes performances sur de très courtes distances (ordre du mètre) [4]. La zone du Upper Mid Band (7 à 30 GHz) est une piste nouvelle qui pourra donner un compromis entre débit et couverture. Enfin les notions de partage dynamique du spectre entre opérateurs, à l'étude en 5G, restent un axe prometteur de l'amélioration de l'efficacité spectrale.

La spatialisation de l'information

Ce terme générique désigne les procédés à la couche physique par lesquels la communication ne se fait plus en mode *broadcast*, c'est-à-dire en inondant l'espace d'ondes radio de manière indiscriminée comme c'est le cas lorsque des antennes classiques sont utilisées, mais de manière dirigée par l'utilisation de faisceaux. La technologie de base dite *multiple-input multiple-output* (MIMO) inventée dans les années 1990 s'est imposée comme l'une des briques fondamentales de la 5G sous la forme du MIMO massif par lequel les antennes relais peuvent être constituées d'un grand nombre (plusieurs dizaines) d'éléments radiofréquence (RF) compacts. La 6G va rendre la communication toujours plus spatialisée et, ainsi, efficace énergétiquement. Étant combinés intelligemment entre eux, les éléments antennaires permettront la formation de faisceaux très discriminants dans le domaine spatial, c'est-à-dire que l'énergie radio générée à l'émetteur se retrouvera concentrée uniquement aux points de l'espace où se trouvent les récepteurs ciblés. Autre nouveauté, la notion de surfaces réfléchissantes reconfigurables (SRR) [5], conceptualisée il y a une dizaine d'années, généralise l'antenne en une notion de surface synthétique et intégrée dont les propriétés en termes de réflexion des ondes électromagnétiques sont contrôlées. Les SRR peuvent alors jouer le rôle d'antenne relais intelligente permettant de conduire les ondes radio vers une cible en contournant les obstacles à la propagation.

L'intelligence artificielle

Face à la complexification grandissante des réseaux de communication sans fil, l'IA est appelée à jouer un rôle crucial dans la conception et l'auto optimisation de ces réseaux [6]. Il existe deux niveaux fondamentaux d'intégration de l'IA dans la 6G : l'intégration au niveau du cœur de réseau et l'intégration au niveau du *edge*, c'est-à-dire les appareillages en bordure de réseau vers lesquels on vient déléster une partie des fonction et calculs afin de prendre des décisions avec un minimum de latence (*edge-AI*).

L'apprentissage automatique dans le cœur de réseau (ou le *cloud*) joue un rôle déjà important pour l'optimisation du réseau, l'allocation des ressources en fonction des classes de services, la maintenance, et la détection d'intrusion ou comportements malveillants. Davantage propre à la 6G, l'utilisation de l'IA sur les appareillages en bout de réseau (*smartphones*, capteurs, robots connectés, antennes relais, etc.) dépendra de l'évolution de leur capacité à intégrer des plateformes de calcul puissantes (GPU). L'*edge-AI* permettra la prise de décision distribuée, locale ou collaborative, d'une manière proche du signal ou des capteurs. Ainsi les appareillages prennent un rôle actif, conférant une puissante intelligence collective au réseau [7].

Le sensing

Cette notion redéfinit la fonction des ondes radio, dont le rôle jusqu'à présent a été confiné au transport de l'information, pour en faire des auxiliaires efficaces de sondage de l'environnement. En effet, en analysant avec les outils d'IA la manière dont les ondes radio sont réfléchies par les obstacles RF, il est possible de récupérer des informations de cartographie précises et des informations sur la présence d'objets (murs, humains, véhicules) aux alentours. Parce que la forme d'onde transmise est typiquement optimisée soit pour le transport de données, soit pour la détection (radar), la 6G introduit le nouveau concept de formes d'ondes hybrides permettant un compromis optimal sous l'appellation ISAC (*integrated sensing and communication*) [8]. À noter que dans certains cas, les ondes peuvent être analysées en des points sélectionnés de l'espace à l'aide de drones intelligents, permettant alors une cartographie et une géolocalisation encore plus précises (voir la Figure 1).

La Communication sémantique

Cette idée revisite les préceptes fondamentaux de la théorie de l'information de Shannon, selon lesquels les fonctions d'encodage sont optimisées pour le transport fiable de données binaires sans égard pour la nature de l'information transportée [9]. Au contraire, le cadre sémantique met en valeur la finalité de la *data*, son degré d'importance dans le temps, et le type d'action menée avec elle au récepteur, pour encoder plus efficacement les infor-



Figure 1 : Station de base cellulaire volante. Le prototype, développé en France par Eurecom, peut assurer une extension de connectivité. Il peut aussi servir de capteur autonome pour la localisation des nœuds terrestres.

mations et ainsi alléger la charge du réseau afin de gagner en capacité ou en efficacité énergétique [10]. Ainsi, par exemple, la communication d'une séquence vidéo à des fins de détection se fera sur la base de l'extraction des éléments clés d'une scène par l'IA, qui seront les seuls à être transmis.

SÉCURITÉ ET IMPACT ÉCOLOGIQUE

Le succès de la 6G ne sera pas mesuré uniquement par sa capacité à offrir de meilleurs tuyaux de données. Comme on a déjà pu l'observer lors du déploiement de la 5G, les enjeux sociétaux sont devenus primordiaux pour l'acceptation de la technologie par le grand public. On prendra ici deux exemples, tout d'abord la question de la sécurité associée à ces systèmes et enfin la question de leur empreinte écologique.

Les aspects de sécurité recouvrent eux-mêmes des questions très diverses. Tout d'abord, la résilience et la redondance des équipements, à renforcer par rapport à la 5G, puis l'intégrité des données véhiculées, la mise en œuvre de protocoles de cryptage et d'authentification robustes, y compris par de nouvelles méthodes issues du quantique, pour se protéger contre les violations de données et les prises de contrôle non-autorisées qui peuvent mettre en danger le fonctionnement de systèmes cyber-physiques essentiels aux infrastructures nationales (usines, centrales nucléaires, aéroports, etc.). Ces menaces seront à spectre plus large que dans la 5G à cause de la multiplication des points de contact dans le réseau [11].

Le deuxième champ d'application est celui de la sécurité individuelle, qui touche à la confidentialité des données et la protection de la vie privée. Dans un mode plus numérisé où chacun est amené à partager une quantité toujours plus grande des données personnelles sur le réseau, la 6G, tout comme les réseaux filaires doivent mettre en place des mécanismes avancés pour la gestion sécurisée des données personnelles.

L'autre grand enjeu sociétal de la 6G concerne son impact environnemental qui recouvre là aussi plusieurs notions diverses allant de l'efficacité énergétique, à l'empreinte globale en passant par les questions de durabilité et recyclabilité des équipements vendus. Avec un accent fort mis dans la 5G sur l'efficacité énergétique dès la phase de *design*, la génération de systèmes de communication en cours a montré un gain en efficacité jusqu'à x 10, en termes de joules par bit transportés, obtenu par une combinaison : de protocoles de désactivation en fonction du trafic, de nouvelles modulations plus efficaces et de densification du réseau. Malgré tout, il est notoire que l'empreinte énergétique globale de la 5G reste supérieure à celle des générations précédentes (même si l'impact très positif que ces systèmes de communication peuvent avoir pour aider d'autres industries à consommer moins, reste globalement insuffisamment mesuré, en particulier à l'échelle mondiale [12] [13] [14]). Le paradoxe s'explique par la multiplication des cas d'usage et la généralisation des applications très haut débit telles que la vidéo VHD sur les *smartphones*. L'amélioration de l'efficacité énergétique reste malgré tout un objectif fondamental du *design* de la 6G, qui pourra se reposer notamment sur les méthodes de spatialisation évoquées plus haut, l'utilisation de l'AI en cœur de réseau pour optimiser les ressources et la mise en œuvre de mécanismes sémantiques qui feront baisser significativement la quantité de données brutes à transporter dans les tuyaux. Tirant les leçons de la 5G néanmoins, il est raisonnable de penser qu'une réduction globale de l'empreinte énergétique ne pourra pas se faire uniquement par des critères d'amélioration techniques du système, et devra faire intervenir des mécanismes incitatifs visant un usage raisonné de ces futurs outils de communication.

CONCLUSION

La 6G cible des objectifs ambitieux en termes de performance et de nouveaux cas d'usage comme la communication immersive et la robotique connectée de précision. De nouvelles briques technologiques sont à l'étude, comme la spatialisation de la transmission, l'IA

intégrée aux fonctions de cœur de réseau ou embarquée dans les appareillages, la communication sémantique, et d'autres avancées encore, qui doivent permettre de répondre aux attentes. Plus largement, un succès humain autour de la 6G ne pourra être assuré qu'en prenant une vue sur les enjeux globaux de ces systèmes avec la prise en compte des attentes du public concernant la sécurité, la protection des droits numériques individuels, et l'impact environnemental.

RÉFÉRENCES

- [1] TANG F., CHEN X., ZHAO M. & KATO N. (2023), "The roadmap of communication and networking in 6G for the Metaverse", *IEEE Wireless Communications*, vol. 30, n°4, pp. 72-81.
- [2] NGUYEN D. C. *et al.* (2022), "6G Internet of Things: a comprehensive survey", *IEEE Internet of Things Journal*, vol. 9, n°1, pp. 359-383.
- [3] WANG C.-X. *et al.* (2023), "On the road to 6G: visions, requirements, key technologies, and testbeds", *IEEE Communications Surveys & Tutorials*, vol. 25, n°2, pp. 905-974.
- [4] TRIPATHI S., SABU N.V., GUPTA A.K. & DHILLON H.S. (2021), "Millimeter-wave and Terahertz spectrum for 6G wireless", In WU Y. *et al.*, *6G Mobile Wireless Networks, Computer Communications and Networks*, Springer.
- [5] LIU Y. *et al.* (2021), "Reconfigurable intelligent surfaces: principles and opportunities", *IEEE Communications Surveys & Tutorials*, vol. 23, n°3, pp. 1546-1577.
- [6] GÜNDÜZ D., DE KERRET P., SIDIROPOULOS N. D., GESBERT D., MURTHY C. R. & VAN DER SCHAAR M. (2019), "Machine Learning in the Air", *IEEE Journal on Selected Areas in Communications*, vol. 37, n°10, pp. 2184-2199.
- [7] XU W., YANG Z., NG D. W. K., LEVORATO M., ELDAR Y. C. & DEBBAH M. (2023), "Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing", *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, n°1, pp. 9-39.
- [8] PIN TAND K. *et al.* (2021), "Integrated sensing and communication in 6G: motivations, use cases, requirements, challenges and future directions", 2021 1st IEEE international online symposium on Joint Communications & Sensing (JC&S), Dresden, Germany.
- [9] LUO X., CHEN H.-H. & GUO Q. (2022), "Semantic communications: overview, open issues, and future research directions", *IEEE Wireless Communications*, vol. 29, n°1, pp. 210-219.
- [10] STRINATI E. C., BARBAROSSA S. (2021), "6G networks: beyond Shannon towards semantic and goal-oriented communications", *Computer Networks*, vol. 190.
- [11] PORAMBAGE P., GÜR G., M. OSORIO D. P., LIYANAGE M., GURTOV A. & YLIANTTILA M. (2021), "The roadmap to 6G security and privacy", *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1094-1122.
- [12] IMOIZE A. L., ADEDEJI O., TANDIYA N. & SHETTY S. (2021), "6G enabled smart infrastructure for sustainable society: opportunities, challenges, and research roadmap", *Sensors*, 21(5), article 1709.
- [13] Numéro « Transitions énergétique et numérique », *Annales des mines - Responsabilité & Environnement*, avril 2023, https://annales.org/re/2023/re_110_avril_2023.html
- [14] FAURE A. & ROUSSILHE G. (2024), Note d'analyse de France Stratégie, « Quelle contribution du numérique à la décarbonation ? », juillet, <https://www.strategie.gouv.fr/publications/contribution-numerique-decarbonation>

Réseaux de communications sur constellations de satellites

Par François BACCELLI
Académie des Sciences

Cet article, qui est issu du rapport de l'Académie des Sciences « Grandes constellations de satellites : enjeux et impacts », a pour objectif de décrire l'organisation, les fonctionnalités et les perspectives des réseaux de communications utilisant des constellations de satellites.

Il se concentre sur les fonctionnalités de communications haut débit. Il décrit les architectures utilisées ainsi que les caractéristiques générales des satellites qui composent ces constellations. Il décrit aussi les principes régissant les communications entre les divers éléments du réseau. Il passe en revue les perspectives à plus long terme de cette classe de réseaux.

INTRODUCTION

Le nouvel âge spatial (*New SpaceX*) marque l'avènement d'une ère nouvelle dans l'utilisation de l'espace caractérisée par une ouverture de l'espace à de nouveaux acteurs fondée sur de nouvelles technologies spatiales, par de nouvelles fonctionnalités pour les satellites mis en orbite et par le développement de constellations de satellites, principalement dans les domaines des communications et de l'observation de la Terre. Ces développements s'appuient sur des avancées scientifiques et technologiques de premier plan ainsi que sur des investissements publics et privés considérables.

Le résultat le plus visible du déploiement de ces nouveaux réseaux spatiaux est de permettre des connexions Internet à haut débit et faible latence en tout point du globe terrestre. Combinés à des capacités d'observation de la Terre, ces nouveaux moyens de communications permettent aussi d'agir en temps réel dans toute région, y compris dans celles dépourvues d'équipements autres que les terminaux.

ARCHITECTURE

L'orbite d'un satellite terrestre s'inscrit dans le plan orbital qui passe par le centre de la Terre et qui est caractérisé par son inclinaison, qui mesure l'angle que ce plan forme avec le plan équatorial, et par le point ascendant de l'orbite. Ce point ascendant est repéré par l'angle, mesuré dans le plan équatorial, entre le point où le satellite traverse le plan équatorial du sud vers le nord et un point donné sur l'équateur. Les orbites circulaires sont définies par un seul paramètre complémentaire qui est l'altitude de l'orbite.

On distingue trois types d'altitudes :

- LEO ou *Low Earth Orbit*, entre 350 et 1 000 km avec une période de l'ordre de la centaine de minutes ;
- MEO ou *Medium Earth Orbit*, typiquement entre 1 000 et 20 000 km, avec une période de quelques heures ;

- GEO ou géostationnaire, avec une altitude de 36 000 km. La période est dans ce cas égale à 24 heures et le satellite reste stationnaire par rapport à la Terre.

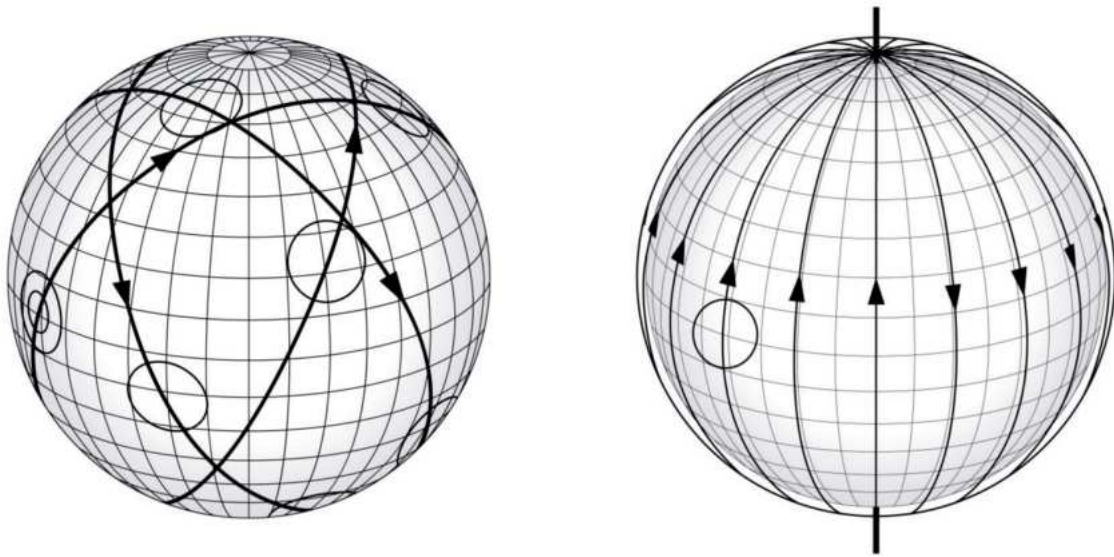


Figure 1 – À gauche : Walker delta (base de Starlink et Galileo) ;
à droite : Walker en étoile (base de OneWeb) (Source : [1]).

La classification de Walker

Deux classes de constellations de satellites, LEO ou MEO, fréquemment utilisées sont les configurations de type Walker delta et Walker en étoile (voir la Figure 1).

Dans une constellation Walker delta, tous les plans orbitaux ont la même inclinaison, les points ascendants de ces orbites sont répartis de manière périodique sur les 360 degrés de l'équateur, les altitudes des orbites sont toutes égales et les satellites d'une même orbite sont positionnés de manière périodique sur cette dernière.

Les constellations Walker en étoile utilisent des orbites avec des inclinaisons proches de 90 degrés (donc quasi polaires). Les plans orbitaux sont organisés de façon périodique sur l'équateur mais sur 180 degrés seulement, si bien que sur l'équateur, les satellites sont tous ascendants sur l'intervalle 0,180 degrés et tous descendants sur le complément.

De nombreuses variantes de ces constellations de base sont utilisées, comme par exemple des combinaisons multi-altitudes, ou star-delta, ou encore LEO-MEO-GEO.

Stations d'ancrage

Les satellites d'une constellation jouent le rôle de passerelle entre des utilisateurs et un réseau/système terrestre (Internet, base de données, etc.). Pour la voie montante, les données émises par l'utilisateur sont envoyées par ce dernier au satellite et renvoyées par le satellite vers des stations au sol qui sont connectées au réseau/système terrestre en question et qu'on appelle des stations d'ancrage.

Pour les orbites basses, en l'absence de communications entre les satellites, il faut beaucoup de stations d'ancrage si on souhaite disposer d'une couverture continue. En effet, la zone de couverture d'un satellite en orbite basse est réduite et la couverture d'un utilisateur requiert que le satellite auquel il est rattaché couvre aussi une station d'ancrage à tout moment. Cette dernière contrainte est levée s'il y a des communications directes entre les satellites permettant de relayer l'information du satellite qui couvre l'utilisateur vers un satellite qui couvre une station d'ancrage. Les communications entre

satellites permettent donc une réduction du nombre des stations d'ancrage et du nombre des satellites.

Satellites

Il y a une grande variabilité dans les caractéristiques des satellites. Il convient en particulier de bien distinguer les satellites utilisés depuis les années 1960 pour les orbites géostationnaires et ceux qui composent aujourd'hui les constellations en orbite LEO ou MEO.

La masse des satellites géostationnaires dépasse plusieurs tonnes (3 tonnes pour Eutelsat Konnect) alors que celle des micro satellites d'observation d'ICEYE est de l'ordre de 70 kg. Starlink s'oriente désormais vers des satellites de deuxième génération plus lourds, d'environ 1 500 kg.

Il faut environ 300 M€ pour la conception et la fabrication d'un satellite géostationnaire. Il faut ajouter à ceci environ 100 M€ pour le lancement et 100 M€ pour la construction de la station d'ancrage. On obtient ainsi un coût global d'environ 500 M€ pour un satellite qui a une durée de vie de 15 à 20 ans. Le coût d'un satellite en orbite basse est bien plus modeste et peut être considéré comme celui d'un consommable. Pour OneWeb, l'ordre de grandeur du coût d'un satellite est de 1 M€.

FONCTIONNALITÉS

Les constellations de satellites sont utilisées dans des contextes multiples qu'on peut classer en trois grandes catégories :

- Les télécommunications, principalement les communications haut débit de type accès Internet universel, mais aussi les communications bas débit pour l'Internet des objets. C'est l'accès Internet haut débit qui conduit aux très grandes constellations actuelles ou en projet qui s'appuient sur des dizaines de milliers de satellites LEO ou MEO.
- L'observation de la Terre : cette fonctionnalité, très diversifiée, conduit au déploiement de nombreuses constellations, chacune avec un nombre modéré de satellites.
- La géolocalisation et la distribution du temps : ce domaine déjà ancien comporte peu de constellations chacune avec un nombre réduit de satellites MEO.

Cet article se concentre sur les communications haut débit. Les très grandes constellations en orbite basse ou moyenne permettent une couverture Internet qui est à la fois universelle (malgré des contraintes en latitude pour certaines configurations), sans interruptions (il faut à cet effet beaucoup de satellites dans le cas LEO), à haut débit (sous des contraintes fortes sur le nombre d'utilisateurs par satellite), avec une latence réduite et avec une bonne efficacité spectrale (grâce aux orbites basses) et résiliente car distribuée.

Les éléments du réseau

Le réseau comporte trois types d'éléments : les terminaux, les stations d'ancrage, qui sont les éléments terrestres, et les satellites en orbite. Les terminaux sont équipés d'antennes qui permettent de communiquer directement avec les satellites. Les satellites communiquent avec les terminaux, éventuellement avec les autres satellites proches et avec les stations d'ancrage. Ces dernières sont des stations au sol qui peuvent être vues comme des premiers points d'accès vers l'Internet et ses bases de données. Le réseau a pour fonction d'interconnecter à tout moment chaque terminal avec une station d'ancrage au moyen d'une connexion comportant au moins un satellite et donc au moins deux liens radio (celui entre le terminal et le satellite et celui entre le satellite et la station d'ancrage).

Connectivité

La connexion entre l'utilisateur final et l'Internet comporte un lien utilisateur-satellite, puis une série de liens inter-satellites et enfin un lien satellite station d'ancrage. S'il n'y a pas de connexions entre satellites, pour avoir une couverture continue, il faut donc qu'à tout moment et pour tout utilisateur, il y ait au moins un satellite ayant la double propriété suivante : être en vue de l'utilisateur et être en vue d'une station d'ancrage.

La situation est différente lorsque les satellites communiquent entre eux. Dans ce cas, au prix de chemins de connexion plus longs, c'est-à-dire avec plus de liens inter-satellites, on peut obtenir un accès haut débit dans une région du globe, tant sur la voie montante que sur la voie descendante, sans avoir à s'appuyer sur des stations d'ancrage dans cette région.

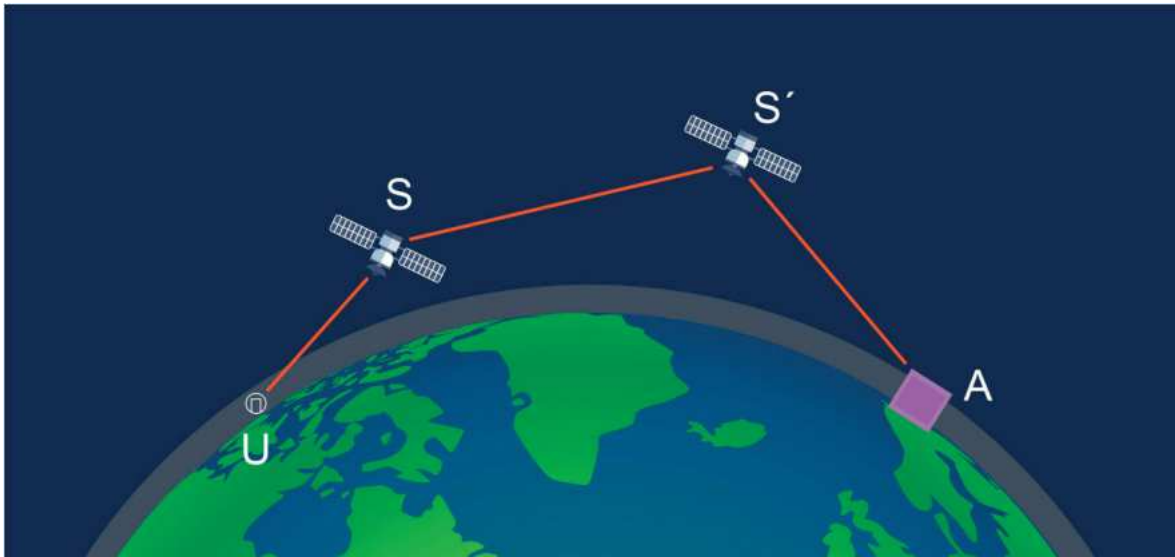


Figure 2 – Les éléments du réseau : U : *utilisateur final*, S et S' : *satellites*, A : *station d'ancrage*. Exemples de lien S-U et S'-A : *Bandes Ku et Ka*.
Liens entre satellites : *Laser* (Source : [1]).

Fréquences

Communications utilisant les radiofréquences (RF)

Dans le contexte du haut débit, les constellations de satellites actuelles et bon nombre des constellations proposées utilisent ou utiliseront les bandes radiofréquences Ku (10,7-14,5 gigahertz [GHz]), Ka (17,3-30 GHz) et V (37-50,4 GHz) pour les liens radio entre les satellites et les éléments de réseau au sol. C'est le cas notamment pour Starlink. Il est aussi envisagé d'utiliser les bandes de la 5G dans le cadre d'accords avec les opérateurs terrestres des réseaux 5G.

Communications optiques et THz

Les communications entre satellites, qui doivent assurer des débits beaucoup plus importants, peuvent se faire en utilisant des faisceaux laser optiques. Cette option est celle des dernières versions des satellites de la constellation Starlink.

Une alternative qui est étudiée est la bande de fréquence dite THz qui couvre la bande (100 GHz, 10 THz). Au sol dans une atmosphère humide, ces bandes de fréquences permettent uniquement des communications sur de courtes distances (quelques km) du

fait de l'atténuation des ondes THz par la vapeur d'eau atmosphérique. Par contre, les ondes THz ne sont pas sensibles à la turbulence atmosphérique.

Antennes

Dans les cas LEO ou MEO, les antennes au sol sont plus complexes que dans le cas géostationnaire car elles doivent pouvoir suivre le satellite assurant le raccordement et changer de satellite pour maintenir la connexion.

Les antennes des stations d'ancrage gèrent de grands débits de données, possèdent une grande ouverture et sont pointées vers les satellites proches.

Les antennes situées à bord des satellites doivent pouvoir former plusieurs milliers de faisceaux simultanés vers les différents utilisateurs. Ce sont des antennes sophistiquées utilisant des réseaux multiples d'éléments répartis sur des surfaces de quelques m² capables d'orienter, en temps réel, de façon adaptative des faisceaux dans la zone géographique concernée.

Quant aux utilisateurs, ils ont, eux aussi, besoin d'antennes capables de suivre les satellites en mouvement et exploitent à cet effet des techniques adaptatives.

Capacité

Dans le contexte des télécommunications, les notions de capacité sont utilisées à divers niveaux. On distingue la capacité globale d'un satellite, la capacité globale d'une constellation ou encore la capacité effective d'un lien radio.

À titre d'exemple, les satellites de Starlink sont présentés comme ayant une capacité globale de l'ordre de 10 Gigabit/s.

Les constellations en cours de déploiement sont conçues pour offrir à chaque utilisateur, dans des conditions idéales, des débits effectifs comparables à ceux de la fibre optique. Par conditions idéales, on entend la proximité du satellite de raccordement, le contrôle des interférences et l'absence de partage des ressources radio entre un trop grand nombre d'utilisateurs.

Latence

Les satellites géostationnaires ont des délais de propagation de 500 millisecondes, ce qui conduit à des latences trop importantes pour certaines applications. Les satellites LEO réduisent ce délai à quelques millisecondes dans les cas les plus favorables. Le contrôle de la latence est essentiel dans le contexte d'applications avec des contraintes de type « temps réel ».

Résilience

Les réseaux à base de constellations LEO et MEO ont de remarquables propriétés de résilience : comme les réseaux à base de satellites géostationnaires, ces réseaux restent fonctionnels même dans le cas de situations locales dégradées, de catastrophes naturelles ou de destruction des réseaux terrestres. La nature distribuée, mobile et reconfigurable des constellations LEO et MEO rend par ailleurs leur attaque physique bien plus difficile que celle des stations au sol ou celle des satellites géostationnaires.

Universalité

Les constellations en étoile couvrent la Terre dans son ensemble y compris les régions polaires. Les constellations delta permettent une couverture en latitudes paramétrable. Celles utilisant des orbites basses ou de moyenne altitude s'appuient sur des densités de satellites élevées et des distances entre satellites réduites pour assurer un service sans interruption.

EXEMPLES

- OneWeb, initialement lancée par le Royaume-Uni, est constituée de 624 satellites configurés en étoile. OneWeb a été rachetée plus récemment par Eutelsat.
- O3b, initialement lancée aux États-Unis par Greg Wyler (un des pionniers du *New Space*) en 2007, constituée actuellement d'une trentaine de satellites produits par Thales Alenia Space en orbite MEO (7 500 km) équatoriale. Cette constellation est détenue désormais par la société SES.
- Starlink (Elon Musk, États-Unis), est la plus grande des constellations. Elle comporte début juillet 2024 environ 5 600 satellites en configuration delta avec des altitudes entre 350 et 500 kilomètres. Elle devrait en comporter 8 000 fin 2024. Elle a une autorisation pour 42 000 satellites. Starlink peut être vue comme la seule constellation en fonctionnement qui offre un accès haut débit quasi-universel à l'Internet et dont le modèle économique soit à l'équilibre.

Voici quelques-uns des projets en cours dans le domaine de l'accès à haut débit à l'Internet :

- Kuiper, constellation d'Amazon, États-Unis, qui comportera 10 000 satellites organisés en plusieurs configurations de type delta.
- IRIS2, projet de l'Union européenne qui comportera quelques centaines de satellites LEO et MEO. Cette constellation est notamment conçue pour garantir un moyen de communication sécurisé pour les gouvernements des pays de l'Union européenne.
- Guowang une constellation de satellites chinoise qui devrait comporter 13 000 satellites.

QUELQUES ORDRES DE GRANDEUR

Selon l'ESO, jusqu'à 100 000 satellites LEO pourraient être lancés en orbite terrestre basse au cours de la prochaine décennie. En supposant que les plans des deux entreprises les plus avancées dans ce domaine soient réalisés, près de 80 000 satellites pourraient être présents à diverses altitudes entre 328 et 1 325 km. En 2022, G. Wyler, fondateur de OneWeb et de O3b, a créé e-space qui prévoit de lancer 100 000 à 300 000 micro satellites en orbite basse pour les besoins de communications et d'observation.

PERSPECTIVES

Internet de l'espace

Plusieurs constellations ont déjà des capacités d'interconnexion des satellites entre eux. Les satellites d'Iridium communiquent sur la bande Ka et les satellites de deuxième génération de Starlink communiquent par laser. Une autre perspective plus générique est celle des communications entre les constellations. La norme "Space Based Adaptive Communications Node", qui rend ceci possible, vient d'être proposée par la DARPA. Ceci permet d'envisager un nouvel Internet de l'espace, infrastructure dynamique composée de satellites multi-orbites et complémentaire du cœur de réseau Internet terrestre.

Les avantages d'un tel réseau sont les suivants : une capacité de projection quasi instantanée sur tout le globe, indépendamment des changements ou attaques qui pourraient avoir lieu sur le réseau terrestre ; une résilience accrue pour le réseau Internet global formé de la conjonction du réseau terrestre et du réseau spatial ; une capacité nouvelle d'observation et de communication conjointes.

Intégration logicielle des réseaux terrestres et satellitaires

Une autre perspective est celle de l'intégration des satellites dans la 5G et la 6G. Cette intégration va demander une refonte de l'organisation des communications satellitaires actuelles et augmenter la dépense énergétique à bord mais aura pour avantage à terme une complète compatibilité avec les protocoles des réseaux cellulaires.

Communications et calcul dans l'espace

Une autre perspective à plus long terme est celle d'un déplacement d'une partie de la puissance de calcul et de stockage vers ces constellations, dans le prolongement de l'idée de localiser le calcul en périphérie de réseau (*edge computing*) dans les satellites de ces constellations. Cette localisation réduit la latence et peut être utile dans le contexte d'applications temps réel.

On peut donc envisager à terme un Internet spatial qui combinera une infrastructure spatiale de routage, du calcul embarqué (notamment pour le traitement du signal et les services temps réel) et des fonctions de communication et d'observation.

ENJEUX

Une première question est celle du contrôle de ces nouvelles classes de réseaux. Les réseaux de communications fondés sur des constellations s'affranchissent de fait des règles (mécanismes d'attribution des fréquences, règles de confidentialité sur les données transmises, règles de localisation des cœurs de réseaux, etc.) qui sont imposées par les États aux opérateurs des réseaux terrestres offrant des services sur leur sol. Dès aujourd'hui, ces réseaux satellitaires peuvent se passer complètement de stations d'ancrage dans les pays qu'ils couvrent. Leur déploiement est donc une perte de souveraineté directe des États sur ce secteur.

Une seconde question est celle du modèle économique de ces nouvelles constellations. Si ces réseaux permettent d'assurer une couverture des zones blanches, ils ne pourront pas remplacer les réseaux actuels pour l'accès à internet. Ainsi, il n'est pas certain que les constellations déployées aujourd'hui puissent devenir économiquement viables à long terme. Les enjeux de souveraineté restent donc les raisons les plus fortes pour leur développement. Ils procurent en effet à ceux qui les contrôlent un moyen de communication haut débit sécurisé à faible latence, un moyen d'observation instantanée d'événements partout dans le monde et donnent donc de nouveaux moyens d'action.

Une troisième question majeure est celle de l'impact de la densification en satellites et en débris sur l'environnement atmosphérique. Dans le contexte de croissance incontrôlée qui prévaut actuellement, l'augmentation du nombre d'objets en orbite fait que les manœuvres d'évitement deviennent de plus en plus fréquentes et elle conduit à une multiplication du nombre de collisions. L'impact des constellations sur l'astronomie est également préoccupant car il touche à la fois les observations optiques et infrarouges et celles qui sont réalisées dans le domaine radio.

L'analyse des impacts négatifs fait apparaître un besoin de régulation internationale d'un domaine qui pour le moment se développe en l'absence de toute contrainte (si ce n'est celle du coût des lancements) et elle souligne la nécessité d'initiatives et d'actions engagées par les parties prenantes pour identifier des principes et des codes de bonnes pratiques qui puissent être adoptés par un nombre croissant de pays et qui permettent d'aboutir rapidement à une régulation internationale du secteur.

RÉFÉRENCE

[1] BACCELLI F., CANDEL S., PERRIN G. & PUGET J-L. (2024), « Grandes constellations de satellites : enjeux et impacts », Rapport de l'Académie des Sciences, mars 2024, <https://hal.science/hal-04607907>

Maîtrise de la complexité dans la 6G - L'approche par la géométrie stochastique

Par François BACCELLI
Académie des Sciences

La 6G est fondée sur l'utilisation de plusieurs nouveaux types d'éléments de réseau, de fréquences et d'architectures. La maîtrise de la complexité de la couverture radio qui en résulte requiert le développement de nouveaux outils de conception et d'analyse. Cet article a pour but de décrire les principes de l'approche de ces questions par la géométrie stochastique ainsi que les perspectives qui se dessinent pour l'utilisation de cette dernière dans le cadre de la 6G.

LA BARRIÈRE DE COMPLEXITÉ DE LA 6G

La 6G est fondée sur l'utilisation de plusieurs nouveaux types d'éléments de réseau comme les relais à base de surfaces intelligentes reconfigurables (SIR) ou les stations de base du réseau non terrestre (RNT), sur l'utilisation de fréquences avec des propriétés de propagation complètement différentes par rapport à la gamme inférieure à 6 GHz utilisée jusqu'à maintenant dans les réseaux cellulaires, sur le développement de nouvelles fonctionnalités fondamentales comme l'intégration des fonctions de communication et la détection (ICD), sur des structures de réseau complètement nouvelles comme les réseaux sans cellule ou les réseaux tridimensionnels.

Si chacune de ces nouveautés majeures fait individuellement l'objet d'études approfondies dans les centres académiques et industriels du domaine, il n'existe aujourd'hui aucun moyen d'évaluer comment tous ces nouveaux éléments et fonctionnalités vont interagir. Que se passe-t-il lors de la densification des relais de type SIR dans un réseau cellulaire ? Pour quels scénarios est-il (spectralement) plus efficace de densifier les stations de base cellulaires plutôt que les SIR ? Quel sera l'effet d'un RNT dense sur le réseau terrestre avec lequel il partage des ressources radio ? Dans le cadre de l'intégration des communications et de la détection, quelle proportion de ressources radio doit être utilisée pour la détection (par exemple celle des obstacles) afin d'optimiser l'efficacité spectrale des communications ? Chacune de ces questions se décline en au moins deux sous-types en fonction de la gamme de fréquence considérée, suivant qu'on utilise les bandes de fréquences en dessous de 6 GHz, avec de bonnes propriétés de propagation, ou les bandes millimétriques dont la propagation est bien plus affectée par les obstacles.

Cet échantillon de questions révèle le besoin d'outils pour la maîtrise de la complexité de l'éventail des possibles en termes de couverture radio, condition *sine qua non* du développement industriel de la 6G dans son ensemble.

Il n'y a actuellement que deux méthodes connues pour ces questions : la simulation à événements discrets et la géométrie stochastique. L'avantage de la simulation est qu'elle ne nécessite aucune hypothèse de modélisation. Ses inconvénients sont connus : elle est basée sur des outils logiciels propriétaires et coûteux ; elle reste par ailleurs inefficace pour l'optimisation. La géométrie stochastique, décrite dans ce qui suit, nécessite au

contraire des hypothèses simplificatrices mais est basée sur des outils mathématiques reproductibles et peu coûteux. Elle est de plus idéale pour l'optimisation. Cet article a pour but de donner les principes de cette dernière approche ainsi que les perspectives qui se dessinent pour son utilisation dans le cadre de la 6G.

GÉOMÉTRIE STOCHASTIQUE ET COMMUNICATIONS

La géométrie stochastique est une branche ancienne des probabilités développée par les écoles des mines européennes dans le contexte de la science des matériaux. L'utilisation de la géométrie stochastique pour les réseaux sans fil est récente [1]. Elle constitue aujourd'hui le principal (et sans doute le seul) outil pour la conception au niveau système de grands réseaux de communications sans fil autre que la simulation. Cette approche est aujourd'hui utilisée au niveau mondial pour la conception et le dimensionnement des réseaux cellulaires. Elle a déjà été utilisée avec succès pour le déploiement des réseaux cellulaires 3, 4 et 5G. Elle est aussi l'outil fondamental pour la conception des réseaux auto-organisés, des réseaux *wifi*, des réseaux véhiculaires, etc.

On peut voir cette approche comme une extension spatiale de l'approche consistant à représenter le trafic Internet et les réseaux de files d'attente dans les routeurs en tant que fonctionnelles de processus ponctuels unidimensionnels (les processus de soumission des paquets IP). Dans ce nouveau contexte, les propriétés du réseau de communications sont vues comme des fonctionnelles de processus ponctuels bidimensionnels ou tridimensionnels exploitant les propriétés statistiques des divers éléments du réseau (stations de base, utilisateurs).

Dans le modèle de base de l'architecture cellulaire, les stations de base (SB) sont disposées selon un processus ponctuel de Poisson homogène d'intensité λ dans le plan euclidien. Les utilisateurs sont aussi disposés selon un processus ponctuel homogène et indépendant. Chaque utilisateur reçoit le signal de la SB la plus proche. Les cellules de raccordement sont donc celles de la tessellation de Voronoï du processus de Poisson des SB (voir la Figure 1 page suivante).

Ce modèle très simple permet de calculer la loi du débit de Shannon d'un utilisateur typique. Considérons par exemple la voie descendante (des SB vers les utilisateurs) dans le cas où on utilise des fréquences différentes pour des utilisateurs différents d'une même cellule. Dans le cas où on considère l'interférence venant des autres cellules comme du bruit, ce débit est proportionnel à

$$\log\left(1 + \frac{S}{I+N}\right)$$

où S est la puissance du signal reçu par l'utilisateur, qui provient de la SB la plus proche, I est la puissance de l'interférence qui provient des stations de base en dehors de la cellule de Voronoï de l'utilisateur en question, et N est la puissance du bruit thermique.

Dans ce modèle, les distributions jointes de S et I sont alors obtenues par des considérations probabilistes assez simples. La distribution de la distance à la SB la plus proche est une loi de Rayleigh comme le montre un argument classique de boule vide. Conditionnellement au fait que la SB la plus proche est à distance r , la puissance de l'interférence est caractérisée simplement par sa transformée de Laplace, qui n'est autre que la fonctionnelle de Laplace d'un processus de Poisson inhomogène, d'intensité 0 dans la boule de centre 0 et de rayon r et d'intensité λ en dehors de cette boule. On peut donc donner une forme explicite pour la distribution de ce débit et donc quantifier l'efficacité spectrale de tels réseaux sous forme close en fonction de quelques paramètres physiques fondamentaux.

Ce modèle, qui a été introduit dans [2], est le plus simple possible dans une classe qui a bien sûr de nombreuses variantes et extensions bien plus réalistes.

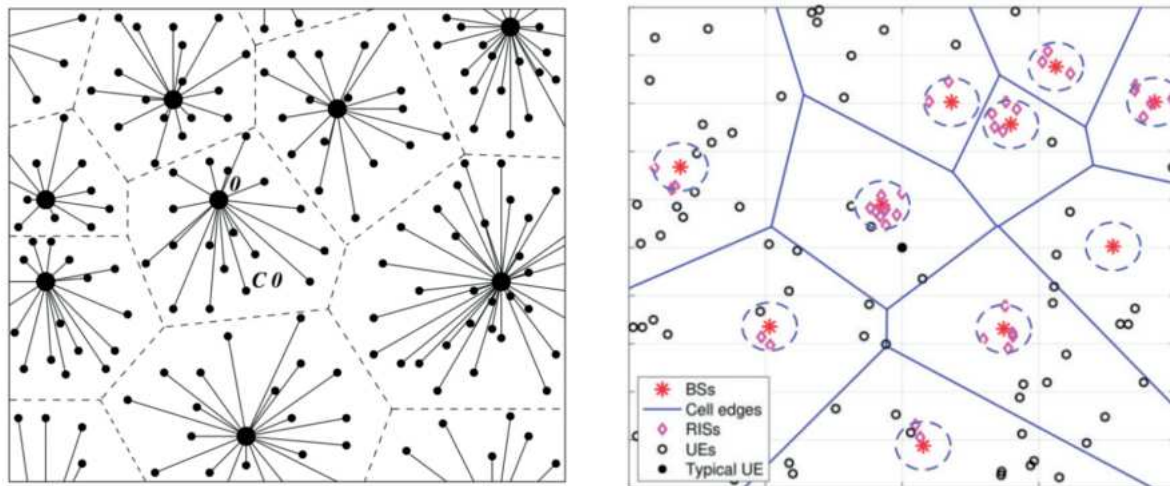


Figure 1 : À gauche : les éléments du réseau cellulaire : stations de base, utilisateurs et liens radios entre les utilisateurs et la station de base la plus proche. La cellule de Voronoï de la SB 0 est notée C_0 . À droite : l'addition d'un nuage de SIR dans le modèle de base (Source : [6] pour la figure de gauche, F. Baccelli et S. Zuyev pour celle de droite).

Le point essentiel est qu'on peut voir cette approche comme une instanciation des analyses de type physique statistique qui permettent de déterminer les moyennes d'ensemble dans un système de particules (ici les éléments de réseau) en interaction (ici les lois de Shannon). En effet, la distribution calculée dans ce modèle est aussi la distribution empirique des débits obtenus sur la voie descendante par les utilisateurs situés dans une grande boule du plan euclidien.

Les résultats de ce type peuvent à leur tour être utilisés pour déterminer les architectures optimisant certaines métriques, par exemple économiques. On trouve un tel exemple de question traité dans [3]. Pour une telle architecture de réseau, quel est l'effet d'une densification des stations de base ? On montre que la densification devient contre-productive au-delà d'un certain seuil : pour des fonctions d'atténuation réalistes, il existe un seuil pour λ au-delà duquel l'efficacité spectrale décroît. Ce fait, qui est dû à une augmentation de la puissance de l'interférence plus grande que celle du signal, a bien sûr d'importantes conséquences économiques.

GÉOMÉTRIE STOCHASTIQUE POUR LA 6G

Les travaux en cours montrent que cette approche permet une évaluation au niveau système de plusieurs paradigmes majeurs de la 6G, tels que les réseaux ICD, les architectures sans cellule, les réseaux basés sur les SIR et les réseaux non terrestres. Ces deux derniers exemples sont brièvement décrits à titre d'illustration dans la suite de cette section.

Réseaux cellulaires et SIR

Le modèle de base (introduit dans [6]) associe un nuage de SIR à chaque SB. Les SB forment toujours un processus ponctuel de Poisson d'intensité λ dans le plan. Dans chaque cellule de Voronoï, on associe à la SB un nuage poissonnien de SIR (ce modèle est connu sous le nom de modèle d'agrégat de Matérn). Ces SIR relaient le signal de la SB vers les utilisateurs de la cellule par focalisation. Ceci est particulièrement utile en cas de blocage entre la SB et l'utilisateur. Dans le modèle de la partie droite de la Figure 1, les SIR forment un processus de Poisson d'intensité constante dans un anneau $[r, R]$ autour de chaque SB.

On peut de nouveau donner des formes explicites pour l'efficacité spectrale de ce modèle et de ses variantes en généralisant l'approche cellulaire de base. Voici quelques questions auxquelles ce modèle permet ou permettra de répondre : quelle est l'influence sur l'efficacité spectrale de la géométrie des nuages de SIR ? Vaut-il mieux peu de SIR avec plus d'éléments réflecteurs contrôlables ou le contraire ? Quel est le gain moyen d'efficacité spectrale apporté par les SIR ? Comment ce gain dépend-il de la densité des obstacles ? Dans une configuration donnée, vaut-il mieux investir dans des SIR ou densifier les SB ?

Réseaux non-terrestres

Les réseaux non-terrestres à base de constellations de satellites en orbites basses ou moyennes se développent fortement partout dans le monde. Ces réseaux seront intégrés dans la 6G. Les stations de base forment maintenant un processus spatial avec des satellites postés sur des orbites avec une certaine inclinaison comme illustré sur la Figure 2. Ces réseaux offrent déjà un accès Internet universel. Ils permettent aussi la collecte de données IoT et l'observation de la Terre. Ils posent aussi de nouvelles questions de souveraineté : nouveaux opérateurs, nouvelles formes de coopération avec les réseaux terrestres, impact sur tout un ensemble de secteurs scientifiques et économiques.

Des premiers modèles à base de géométrie stochastique sphérique sont en cours de développement [4, 5], basés sur la notion de processus d'orbites et de processus ponctuels sur ces orbites. Les notions de couverture, de raccordement au plus proche ont pu être étendues et la distribution de Shannon a pu être établie pour quelques exemples simples. De très nombreuses questions se posent au niveau du système : des questions techniques sur les débits atteignables, les *handovers*, l'ordonnancement et le routage. Des questions économiques sur les interactions et interférences entre les réseaux terrestres 5G et ces réseaux non terrestres. Ou encore des questions sur le rôle des SIR pour la connectivité avec le réseau non terrestre [7].

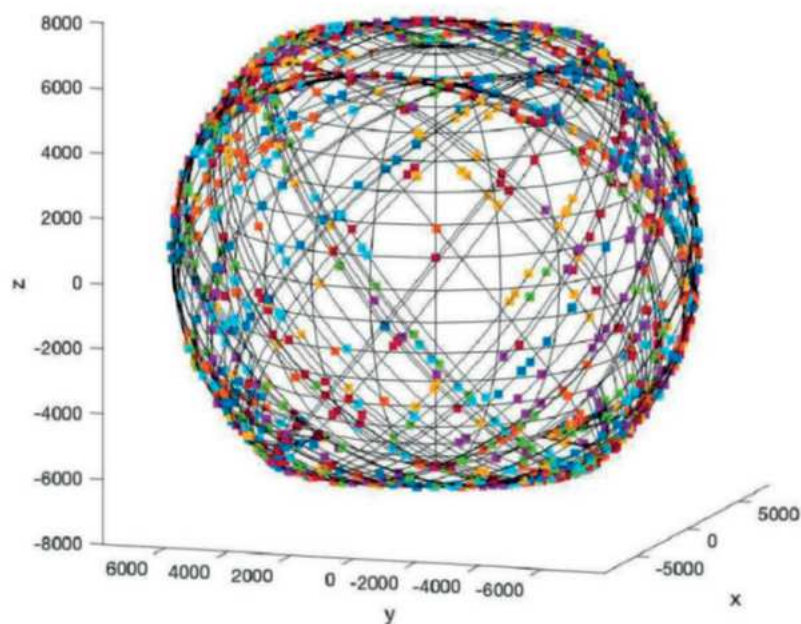


Figure 2 : Une constellation de satellites LEO : Géométrie sphérique avec SB en orbite. Les SB se déplacent sur des orbites avec une inclinaison fixée (Source : <https://arxiv.org/abs/2212.03549>).

Au-delà de ces exemples, le point essentiel est que cette approche par la géométrie stochastique permet aussi d'analyser les interactions entre les divers paradigmes décrits dans les sections précédentes. Par interaction, nous entendons ici une gamme d'activités allant de l'analyse économique conjointe à l'optimisation conjointe du déploiement. Elle permet aussi la conception de nouveaux contrôleurs temps réel et non temps réel pour tous ces nouveaux paradigmes.

Il existe une forte communauté de chercheurs de haut niveau travaillant sur ces questions en France, où cette approche est née. Si elle est développée et structurée correctement, une nouvelle phase de géométrie stochastique sans fil pour la 6G pourrait permettre à l'Europe de construire un outil pour évaluer et optimiser l'intégration de tous ces nouveaux paradigmes en 6G.

RÉFÉRENCES

- [1] BACCELLI F. & BLASZCZYSZYN B. (2009), "Spatial modeling of wireless communications – a stochastic geometry approach", *Foundations and Trends in Networking*, NOW Publishers.
- [2] ANDREWS J., BACCELLI F. & GANTI R. (2011), "A tractable approach to cellular network modeling", *IEEE Transactions on Communications*, 59(11).
- [3] ALAMMOURI A., ANDREWS J. G. & BACCELLI F. (2019), "A unified asymptotic analysis of area spectral efficiency in ultradense cellular networks", *IEEE Trans. Information Theory*, 65(2).
- [4] OKATI N., RIIHONEN T., KORPI D., ANGERVUORI I. & WICHMAN R. (2020), "Downlink coverage and rate analysis of low earth orbit satellite constellations using stochastic geometry", *IEEE Trans. Commun.*, 68(8).
- [5] CHOI C.S. & BACCELLI F. (2024), "Cox point processes for multi-altitude LEO satellite networks", *IEEE Trans. Veh. Technol.*, à paraître.
- [6] SUN G., BACCELLI F., FENG K., UZEDA GARCIA L. & PARIS S. (2024), "A stochastic geometry framework for performance analysis of RIS-assisted OFDM cellular networks", preprint, <https://arxiv.org/abs/2310.06754>
- [7] LEE J. & BACCELLI F. (2024), "How much can reconfigurable intelligent surfaces augment sky visibility: a stochastic geometry approach", preprint, <https://arxiv.org/abs/2403.08930>

Systèmes cyber-physiques et jumeaux numériques, déclencheurs et conditions d'une adoption accélérée

Par Pascal BROSSET
Accenture

Le terme de systèmes cyber-physiques (SCP), apparu à la même époque que l'Industrie 4.0, a été l'objet d'une abondante littérature académique en décrivant le potentiel dans des domaines d'application allant de l'industrie à la santé. Pour rappel, le terme a été introduit en 2006 par le Dr Helen Gill de la National Science Foundation (NSF) aux États-Unis. Ce concept a été développé pour décrire l'intégration de l'informatique avec les processus physiques, où des systèmes embarqués et des réseaux surveillent et contrôlent les processus physiques, généralement avec des boucles de rétroaction où les processus physiques affectent les calculs et *vice versa*.

La convergence de facteurs socio-économiques et du progrès en termes d'intelligence artificielle crée une accélération dans l'adoption des SCP, souvent confondus avec les jumeaux numériques. Cet article explore, du point de vue des industriels, les raisons de cette accélération et ses implications.

L'un étant le pendant virtuel de l'autre, les jumeaux numériques et les systèmes cyber-physiques sont indissociables, surtout dans le domaine des opérations. C'est pourquoi le deuxième terme est aujourd'hui beaucoup plus utilisé par les industriels et nous l'utiliserons aussi dans cet article qui adopte résolument ce point de vue pratique.

Encore relégué au rang de concept il y a encore moins de 10 ans, avec autant d'interprétations que de directeurs industriels, les jumeaux numériques sont maintenant acceptés comme la fondation conceptuelle et technique de toute stratégie ambitieuse de digitalisation des opérations dans l'industrie. Un jumeau numérique opérationnel est aujourd'hui généralement compris comme un modèle (sémantique) complet des opérations en représentant tant la dimension physique que procédé. Ce modèle permet d'évaluer en quasi temps réel tout événement opérationnel avéré ou anticipé, allant de la dérive d'un paramètre de processus à l'annulation d'une commande, par rapport à l'état de l'ensemble des opérations. Cette évaluation contextualisée permet la décision, humaine ou automatisée d'actions correctives minimisant l'impact sur la performance. La généralisation des jumeaux numériques opérationnels permet donc d'envisager des usines dites autonomes, c'est-à-dire capables d'anticiper et de s'adapter en continu et de façon optimale aux inévitables variations de leur environnement. C'est pourquoi McKinsey estime que le lien entre les mondes physique et numérique, facilité par les jumeaux numériques, pourrait générer jusqu'à 11,1 trillions USD par an en valeur économique d'ici 2025. Cet impact inclut des secteurs tels que l'aérospatiale, la défense, l'industrie manufacturière et les biens de consommation courante.

Dans cet article, nous proposons d'explorer les raisons de cette évolution et de son accélération récente, ses implications pour l'industrie afin que cette accélération génère effectivement les résultats escomptés.

FACTEURS D'ACCÉLÉRATION SOCIO-ÉCONOMIQUES

Rareté des talents

Suite au changement de génération en cours dans l'industrie, et pour des raisons sans doute liées au système éducatif, les industriels font face à une difficulté croissante et devenant critique à recruter de nouveaux talents, des opérateurs aux ingénieurs de production. L'expertise tacite qui, encore aujourd'hui « fait tourner » les usines n'est plus transmise et doit être formalisée et même modélisée. Par leur capacité à intégrer des données hétérogènes pour modéliser les opérations et intégrer des logiques d'apprentissage, les jumeaux numériques permettent cette formalisation.

Corollairement, la notion d'usine autonome¹, jusqu'alors objectif théorique et lointain, devient donc un impératif à moyen terme pour de nombreux responsables industriels dans la plupart des secteurs industriels.

Montée des exigences

Entre le niveau d'automatisation des usines et le déploiement de programmes d'excellence industrielle, les *leaders* des différentes industries ont atteint des niveaux de performance élevés au niveau de chaque équipement ou d'ilots de fabrication. Les prochaines opportunités sont généralement liées à des problèmes de type qualité, économies d'énergie ou adaptation à la demande qui demandent de maîtriser les opérations dans leur ensemble. Là encore, la puissance et flexibilité des jumeaux numériques, liées comme nous le verrons à l'essor de l'IA, les rendent incontournables pour l'optimisation multicritères d'une usine intégrée.

FACTEURS D'ACCÉLÉRATION TECHNOLOGIQUE

La concurrence de ces exigences nouvelles et l'arrivée à maturité de technologies sous-jacentes créent les conditions d'une adoption accélérée des jumeaux numériques dans les 5-10 prochaines années.

Adoption croissante de l'ingénierie systèmes

Une des conditions à la généralisation des jumeaux numériques est, de façon évidente, la capacité à les modéliser, et ce de façon modulaire. La montée en puissance de l'ingénierie systèmes (*model-based system engineering*²) initialement liée à la complexité croissante

¹ Une usine autonome est une installation de production où les processus de fabrication sont largement automatisés et contrôlés par des systèmes cyber-physiques sans intervention humaine directe. Ces usines utilisent des technologies avancées telles que l'Intelligence artificielle (IA), l'Internet des objets (IoT), la robotique, l'apprentissage automatique et les jumeaux numériques pour optimiser les opérations et la prise de décision en temps réel.

² Le *model-based systems engineering* (MBSE) est une méthode d'ingénierie qui repose sur la création et l'utilisation de modèles numériques tout au long du cycle de vie du système pour soutenir les activités de conception, d'analyse, de vérification et de validation. Contrairement aux méthodes traditionnelles basées sur des documents, la MBSE utilise des représentations formelles pour améliorer la compréhension, la communication et la gestion des systèmes complexes.

des produits et systèmes (mécanique > électronique > numérique > connecté) a naturellement conduit à son extension vers la conception des moyens de production. Les suites intégrées d'éditeurs tels que Dassault Systems permettent aujourd'hui la conception, simulation et validation concurrente du produit et de son système de production.

Les systèmes de production modernes, généralement adéquatement instrumentés, sont donc conçus comme ces SCP.

Évolution des architectures industrielles

La deuxième condition de généralisation est la possibilité, non seulement d'acquérir et intégrer un grand nombre de paramètres, mais également de les analyser en temps réel pour détecter les variations correspondant à des événements significatifs par rapport au bruit de fond, et finalement de permettre leur évaluation instantanée dans le contexte du jumeau numérique.

Ce besoin combiné d'instantanéité et de puissance est aujourd'hui bien supporté par les architectures hybrides de type *cloud and edge*³, combinant une capacité de calcul locale (*edge*) et la puissance du *cloud* dans une architecture basée événements à faible latence. À titre d'illustration, un événement généré par le premier niveau d'analyse sur un serveur local peut être évalué dans le *cloud* et déclencher un changement de consigne transmis au système de contrôle en moins d'une seconde.

Il ne s'agit bien entendu pas de remplacer ledit système de contrôle, mais de le compléter pour adresser des dérives, par exemple de qualité, qui demandent à être résolues dans un contexte bien plus vaste que celui de la machine sous le contrôle dudit système.

Pénétration croissante de l'IA et *machine learning*

L'évaluation et résolution de dérives dans un contexte étendu à l'ensemble de l'usine se prête particulièrement bien au *machine learning* (ML⁴), pour peu que les données nécessaires soient disponibles de façon fiable et structurée. Pour le développement de logiques d'évaluation, il est typique de passer 80 % du temps et des ressources à structurer les données et 20 % au développement algorithmique.

Par la rigueur qu'ils amènent naturellement dans la gestion et structuration des données industrielles, les jumeaux numériques accélèrent le développement d'algorithmes de ML qui sont progressivement intégrés dans ceux-ci pour la détection et prédiction d'anomalies et le support à la décision. Cette intégration confère aux jumeaux numériques des capacités d'auto-apprentissage, d'adaptation à l'environnement et d'amélioration de performance.

Jumeaux numériques et IA sont donc indissociables et c'est leur association qui accélère, du point de vue technique, leur adoption dans l'industrie. L'émergence de l'IA générative va amplifier le phénomène.

³ *L'edge computing* consiste à décentraliser le traitement des données, en rapprochant les capacités de calcul et de stockage des dispositifs générant les données, tels que les capteurs IoT, les appareils mobiles, ou les systèmes embarqués. Cette approche réduit la latence, améliore la bande passante et permet une prise de décision plus rapide en analysant les données localement.

⁴ Le *machine learning* (apprentissage automatique) est une sous-discipline de l'intelligence artificielle (IA) qui se concentre sur le développement de systèmes capables d'apprendre et de s'améliorer automatiquement à partir de l'expérience et des données sans être explicitement programmés pour chaque tâche.

Interface en langage naturel

Même s'il est grandement facilité par la structure en graphe, l'accès à l'expertise formalisée par un jumeau numérique peut rester complexe pour l'utilisateur moyen. Cette même structure permet par contre d'utiliser les *large language models* (LLM⁵) pour accéder en langage naturel à cette expertise. Cette capacité est aujourd'hui démontrable et en cours de validation par certains industriels.

Ces mêmes LLMs permettent par ailleurs d'enrichir le jumeau numérique avec la vaste base de documents non structurés, rapports, manuels et même photos, qui sont disponibles mais mal exploités dans les intranets des entreprises. À terme, même si cet usage n'en est encore qu'au stade de l'expérimentation, ce mode conversationnel d'accès permettra également de formaliser l'expertise tacite d'experts et d'en personnaliser l'accès en fonction du niveau d'expertise de l'opérateur y faisant appel.

Prototypage rapide

De même que l'IA générative permet de générer automatiquement des programmes informatiques, elle permet aussi d'accélérer le développement de cas d'usage basés sur le jumeau numérique. Une application simple, déjà disponible aujourd'hui, permet de générer des rapports complexes grâce à des interrogations successives, puis de transformer ces rapports en code exécutable qui sera déclenché en réaction à un événement, fournissant ainsi instantanément le contexte nécessaire à la prise de décision afférente.

Une autre utilisation, de nouveau déjà explorée pour le développement d'applications classiques, est la génération automatique de tests et jeux de données de validation, phase essentielle du déploiement opérationnel de cas d'usage pouvant impacter la sécurité, qualité ou performance des opérations. Outre la validation, cette même capacité de génération de jeux de données synthétiques est utilisée pour l'entraînement d'algorithmes de ML, par exemple pour les systèmes de vision artificielle de plus en plus utilisés pour le contrôle en ligne de procédés complexes.

Capacités de simulation

L'IA générative est utilisée aujourd'hui pour optimiser la conception de produits, par génération de milliers de prototypes, et cette approche peut aussi être utilisée pour simuler différentes configurations de machines/lignes/usines et flux de produits pour identifier les paramètres qui minimisent les temps d'arrêt et maximisent l'efficacité.

Environnements synthétiques de simulation

Une autre conséquence de la difficulté à recruter des opérateurs est un intérêt grandissant pour les robots homomorphiques capables de remplacer ceux-ci sans modification de l'environnement de travail. Ceci demandera une capacité d'apprentissage non seulement des modes opératoires, mais également de l'insertion dans un environnement non dédié. Cet apprentissage ne sera possible que dans des environnements virtuels/synthétiques qui seront aussi utilisés pour augmenter la définition et fidélité de simulations mettant en jeu, par exemple, une automatisation poussée des flux physiques par des flottes de chariots autonomes.

Plus proche de nous, les simulations générées par IA sur la base d'un jumeau numérique peuvent être utilisées pour former les opérateurs dans un environnement virtuel, sans risque pour leur sécurité. Ces simulations peuvent reproduire des scénarios dangereux

⁵ Un *large language model* (LLM) est un type avancé de modèle d'apprentissage automatique spécifiquement conçu pour comprendre et générer du langage naturel. Ces modèles sont caractérisés par leur grande taille, mesurée par le nombre de paramètres qu'ils contiennent, et leur capacité à traiter de grandes quantités de texte pour réaliser diverses tâches linguistiques.

ou complexes, permettant aux travailleurs d'acquérir de l'expérience et des compétences en toute sécurité.

Moteurs génériques de simulation

Ces mêmes capacités de simulation utilisées pour la conception des lignes de production doivent aussi être utilisables lors du fonctionnement de l'usine, par exemple pour reconfigurer les flux internes suite à un changement du planning de production. Les moteurs actuels reposant sur des algorithmes de type événements discrets seront complétés dans certains cas par la même approche de génération et évaluation de scénarios. Cette approche sera grandement facilitée par des moteurs génériques qui peuvent être combinés dans un des environnements de simulation susmentionnés.

Modèles cognitifs et systèmes adaptatifs

Comme mentionné ci-dessus, les jumeaux numériques permettront la formalisation progressive de savoir-faire tacites de gestion des opérations. Qu'en est-il des opérations manuelles qui resteront impossibles à automatiser, comme les assemblages complexes en petite série ?

L'IA générative doit permettre d'analyser les opérations, par exemple *via* des caméras analysant les mouvements des opérateurs, en extraire des séquences type et intervenir ensuite de façon ponctuelle pour guider ou valider l'assemblage des prochains produits du même type. Cette technique est d'ailleurs envisagée pour réduire les risques en salle d'opération en levant potentiellement des alarmes à destination du chirurgien.

Cette approche pose bien entendu des questions importantes de confidentialité mais, moyennant l'explication transparente des mécanismes d'anonymisation mis en place, ces mentors virtuels peuvent se généraliser.

IMPLICATIONS POUR L'INDUSTRIE

Les investissements consentis ces cinq dernières années par les géants du cloud et leurs partenaires font que la technologie est largement disponible pour supporter l'évolution vers des architectures industrielles tirant pleinement parti de jumeaux numériques. Par contre, ceci n'est pas une condition suffisante pour que l'adoption soit généralisée et génératrice de valeur. Cette adoption généralisée demandera des approches collaboratives tant au niveau des secteurs industriels que des entreprises elles-mêmes.

Émergence de standards d'interopérabilité et collaboration

La création d'un jumeau numérique complet d'une usine demande d'agréger un volume considérable de données venant de dizaines de sources internes et externes à l'entreprise. Il est donc indispensable qu'émergent des standards de définition et d'interopérabilité de jumeaux numériques. À terme, tout équipement devra être fourni avec son jumeau numérique prêt à être intégré dans celui de la ligne et de l'usine.

C'est l'objectif du standard *Asset Administration Shell* (AAS) proposé par l'Industrial Digital Twin Association (IDTA), en collaboration avec d'autres organismes normatifs, visant à standardiser la manière dont les informations sont gérées et échangées entre les actifs (machines, dispositifs, systèmes) dans un environnement industriel. Les objectifs de l'AAS sont parfaitement alignés avec les conditions de la généralisation de l'utilisation des jumeaux numériques :

- interopérabilité : faciliter l'échange d'informations entre différents systèmes et équipements en utilisant des standards ouverts et assurer que les actifs de différents fabricants puissent interagir et communiquer de manière transparente ;

- gestion du cycle de vie : gérer et documenter toutes les phases de vie de l'actif, de sa conception à son démantèlement et conserver les données historiques pour faciliter l'analyse et l'amélioration continue ;
- flexibilité et adaptabilité : permettre la configuration et la reconfiguration des actifs en fonction des besoins changeants du processus de production et s'adapter à différentes échelles, des petits dispositifs aux grandes installations industrielles ;
- sécurité et conformité : garantir la confidentialité, l'intégrité et la disponibilité des informations échangées et assurer la conformité avec les normes et réglementations industrielles et de sécurité.

L'émergence de plateformes ouvertes permettant la conception collaborative de jumeaux numériques serait un puissant accélérateur d'adoption de standards tels que l'AAS.

Évolution des organisations

La mise en œuvre d'un jumeau numérique est très différente de celle d'une application de type MES (*manufacturing execution system*⁶) et demande une évolution profonde de la collaboration entre l'informatique, l'*engineering* et les opérations. En effet, un jumeau numérique est en évolution constante, tant du point de vue du modèle de données que des cas d'usage qu'il supporte et s'apparente plus à une plateforme qu'à une application classique. Alors qu'un MES est généralement développé sur la base de spécifications fonctionnelles établies en collaboration entre les utilisateurs et l'informatique, puis modélisée, développée et déployée par l'informatique, un jumeau numérique implique une répartition très différente de rôles :

- le modèle de données, généralement de la responsabilité de l'informatique et conçu selon des critères techniques, devient la responsabilité des opérations et est développé selon des critères opérationnels. L'objectif est que ce modèle de données soit exploitable par des experts, humains ou systèmes, à des fins de décision, et qu'il puisse être conçu et enrichi de façon évolutive. Le rôle de l'informatique est de mettre à disposition la plateforme technologique et méthodologique permettant aux fonctions *engineering* et opérations de collaborer pour la définition, la gestion et la gouvernance du modèle de données et services associés ;
- sa gouvernance, en particulier, doit être assurée par l'*engineering* et les opérations, et non par l'informatique. Cela demande l'émergence de nouveaux rôles telles que « propriétaire de données », en charge de la structuration et qualité des données pour un domaine particulier. L'expérience prouve que ce sont ces rôles qui conditionnent le succès d'une démarche de jumeau numérique ;
- la mise à disposition des services d'accès aux données est aussi fondamentale, la complexité d'un jumeau numérique réservant son accès direct aux spécialistes. Une partie importante de la collaboration entre l'*engineering*, l'informatique et les opérations est donc la définition de services d'accès aux données structurées par le jumeau sous forme de services (*data services/products*) formellement décrits dans un catalogue avec leurs méthodes d'accès. Ces services évolueront de façon régulière, reflétant la couverture croissante du jumeau numérique et de l'intelligence associée.

⁶ Un *manufacturing execution system* (MES) est un système informatique utilisé en milieu industriel pour suivre et documenter la transformation des matières premières en produits finis. Le MES assure la gestion des processus de production, la visibilité et traçabilité des opérations de fabrication et la génération de rapports.

La mise en œuvre de ces évolutions d'organisation et de la gouvernance associée permet une structuration et accélération de l'innovation des procédés et de la valeur associée :

- chaque nouveau cas d'usage augmente la couverture du jumeau numérique et l'intelligence des services associés ;
- ceux-ci sont mis à disposition de l'ensemble de l'organisation sous forme de services intelligents, base d'un prochain cycle d'innovation.

C'est le déclenchement et l'amplification de ce cycle d'innovation qui est le principal objectif de la mise en œuvre d'un jumeau numérique, en support de l'évolution vers l'usine autonome.

CONCLUSION

Les jumeaux numériques sont indissociables et symbiotiques de l'application à l'échelle de l'IA aux opérations industrielles, laquelle permettra de concilier efficacité économique et sobriété environnementale. Les conditions techniques sont réunies pour leur adoption progressive, mais celle-ci demandera des modèles de collaboration différents et plus ouverts, tant dans l'entreprise qu'entre elles.

Confrontée au besoin de retrouver une souveraineté industrielle dans un environnement géopolitique marqué par la montée des dangers, l'Europe a une opportunité unique de se réindustrialiser en tirant pleinement profit de cette digitalisation des *assets*. Avec des champions comme Dassault Systems, Siemens, SAP et Schneider Electric, ces investissements peuvent combiner efficacité économique et sobriété environnementale, tout en offrant des emplois industriels hautement qualifiés et attractifs pour les générations futures.

Les apports de Gaia-X

Par Anne-Sophie TAILLANDIER

Directrice générale de TeraLab, filiale de l'IMT

Et Pierre GRONLIER

Chief Innovation Officer de Gaia-X

Gaia-X, créée en 2020, par les gouvernements allemand et français, vise à créer une infrastructure de données fédérée garantissant la souveraineté et la transparence des données, soutenant ainsi l'économie numérique de l'Europe. Avec plus de 300 membres, Gaia-X promeut l'interopérabilité et la sécurité grâce à un cadre normalisé, facilitant le partage des données tout en maintenant le contrôle et la conformité avec les réglementations européennes telles que le RGPD. Elle met l'accent sur l'autonomie technique, opérationnelle et juridique, en encourageant les stratégies multi-*cloud* afin de réduire la dépendance à l'égard des fournisseurs non européens. L'architecture de Gaia-X s'appuie sur des vocabulaires normalisés, des références vérifiables chiffrées et un registre de confiance pour garantir des échanges de données sécurisés et vérifiables. L'initiative soutient également le développement de l'IA en garantissant la traçabilité des données et le consentement, conformément à des réglementations telles que l'IA Act.

À l'avenir, Gaia-X vise à intégrer les réglementations futures, en fournissant une couverture complète pour les fournisseurs de données et les consommateurs, favorisant ainsi un écosystème numérique digne de confiance.

INTRODUCTION

L'intelligence artificielle générative bouleverse les enjeux des services numériques. Le développement de l'IA a connu une accélération spectaculaire au cours des derniers mois, son impact annuel sur l'économie mondiale est estimé entre 2,6 et 4,4 mille milliards de dollars (étude McKinsey, juin 2023).

Les intelligences artificielles sont issues de trois piliers que sont : les modèles, les données et les services de calcul. Les champions de l'IA générative sont ceux qui ont accès à ces trois composantes.

Depuis 2022, l'Europe a voté un certain nombre de réglementations, Data Act, Data Governance Act, AI Act, Digital Service Act, Digital Market Act, pour permettre le développement d'une économie numérique européenne porteuse de ses valeurs de liberté, de protection des citoyens et d'autonomie stratégique. Et pourtant, plus de 70 % du marché est détenu par les géants du numérique que sont Microsoft Azure, Amazon Web Services, ou encore Google Cloud, rejoints depuis peu par des acteurs venant d'Asie, tels que Huawei, Alibaba et Tencent.

En 5 ans, de 2017 à 2022, la part de marché des acteurs européens est passée de 26 à 13 %, alors que le marché du *cloud* en Europe a lui été multiplié par 5 passant de 2 milliards en 2017 à 10 milliards en 2022¹.

La facilité d'usage de ces services cache souvent un enfermement de l'utilisateur et une non maîtrise des risques. On considère aussi que 80 % des données sont dans les entreprises et partiellement exploitées. Les nouveaux produits et services des entreprises dépendent de leurs données mais aussi de données provenant d'autres entreprises.

Les enjeux de Gaia-X sont la transparence des informations, l'interopérabilité, la vérification des autorisations pour permettre un développement de cette économie équitable où l'utilisateur peut reprendre le contrôle de ses informations

Les objectifs de Gaia-X

Qu'est-ce que Gaia-X

Gaia-X est une association créée en 2021, AISBL, internationale basée à Bruxelles. Elle a été créée à l'initiative des gouvernements allemand et français. Elle a maintenant 300 membres et une vingtaine de Hub Nationaux. Le conseil d'administration est réservé aux entités juridiques dont le siège social mondial est européen.

Son objectif est de permettre une meilleure circulation des données privées en fournissant un cadre à la fois technique et juridique permettant aux fournisseurs de données d'avoir les garanties sur leur protection et leur utilisation. De ces espaces de données pourront apparaître de nouveaux modèles économiques pour les entreprises en travaillant en écosystème. C'est bien dans l'objectif de permettre le développement d'une intelligence artificielle puissante, basée sur des services numériques interopérables et permettant aux utilisateurs de choisir le niveau de sécurité correspondant à la sensibilité de leurs données, mais aussi qui respecte les conditions d'utilisations des données telles que définies par les propriétaires de ces données.

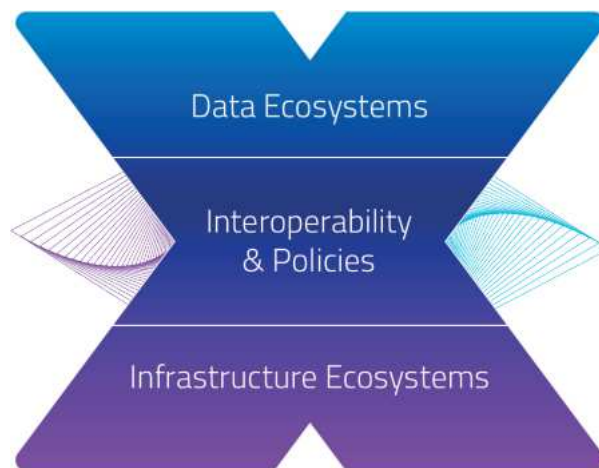


Figure 1 : Gaia-X concerne les écosystèmes d'infrastructure et de données (Source : Gaia-X – Hub France).

L'initiative doit permettre de porter les valeurs européennes de transparence, d'auto-détermination, de protection des données, de sécurité et de portabilité, afin de permettre

¹ <https://www.srgresearch.com/articles/european-cloud-providers-continue-to-grow-but-still-lose-market-share>

aux utilisateurs européens de *cloud* de reprendre leur destin en main. Gaia-X s'appuie sur deux axes majeurs :

- une architecture technique, permettant une décentralisation des identités, des ontologies permettant de décrire les différents objets, les protocoles d'échanges, des nœuds de confiance appelés des *clearing house* ;
- une objectivation de la confiance avec le document de compliance. Les critères de confiance qui y sont décrits couvrent aussi bien les parties contractuelles et réglementaires (comme le RGPD) mais aussi la portabilité, la perméabilité par rapport à des lois extraterritoriales non européennes.

Toutes ces règles et spécifications techniques ont pour objectif de faciliter la circulation des données, dans le but de produire des intelligences artificielles porteuses des valeurs européennes.

L'association est internationale, ses membres, acteurs du numérique et utilisateurs peuvent au sein des groupes de travail de l'association, définir les spécifications techniques et règles de compliance. Elles sont ensuite validées par le conseil d'administration de Gaia-X qui est réservé aux entreprises et institutions dont le siège social mondial est européen. Depuis 2023, Catherine Jestin, CIO d'Airbus, est présidente de l'association. Le Hub France de Gaia-X² est coordonné par l'Institut Mines Télécom.

La confiance

La confiance dans le numérique est une mesure du risque basée sur plusieurs éléments :

- l'évaluation de l'autonomie présente et future, vis-à-vis des services numériques utilisés ;
- le niveau de transparence des services numériques utilisés par rapport aux exigences de conformité, qu'elles soient réglementaires, liées à un domaine métier ou simplement organisationnelles.

La notion de souveraineté est volontairement écartée car soit sans consensus sur ce qui devrait ou ne ferait pas partie du périmètre d'une telle offre de service, soit basée sur l'étymologie du mot et limitée à un territoire géographique sous le contrôle d'une autorité suprême ce qui nous semble incompatible avec les objectifs de la section précédente.

Gaia-X préconise l'utilisation du terme d'autonomie et promeut une évaluation de l'autonomie basée sur trois axes :

- L'autonomie technique : quel est le niveau de répliquabilité, portabilité (total, partiel, nul) de la totalité des couches logicielles et matérielles de l'offre de service ?
- L'autonomie opérationnelle : dans le cas d'une migration d'un service, par exemple en ligne (*cloud*) vers un service sur site (*on premise*), quel est le niveau de ressource et de savoir-faire nécessaire pour maintenir le service en condition opérationnelle ?
- L'autonomie légale : quelles sont les potentielles juridictions applicables au service et aux données stockées, transférées ou traitées par ledit service ?

Le second aspect est celui de l'évaluation de la conformité qui consiste en l'évaluation de critères décrits dans des schémas de conformité.

Ces critères doivent être mesurables, reproductibles, comparables et sont évalués par des acteurs impartiaux accrédités par des autorités reconnues du domaine d'évaluation ou identifiés par l'auteur du schéma de conformité.

² <https://www.gaia-x-hub.fr>

Ces schémas de conformité servent de référence pour identifier des services ou des fournisseurs de services qui répondent à des exigences standardisées, telles que l'hébergement de données de santé, le traitement et stockage de données financières, la cybersécurité, l'environnement, etc.

Gaia-X fournit une méthodologie et des outils prêts à l'emploi qui permettent l'évaluation de l'autonomie et l'évaluation de la conformité.

Il est à noter que la confiance n'est pas automatiquement réciproque. Ces outils sont aussi bien utilisés par les utilisateurs des services et données pour valider et vérifier les engagements du fournisseur, que par les fournisseurs de services et données pour valider et vérifier les engagements des utilisateurs.

Cette unique combinaison permet d'identifier des offres, parmi différents *cloud providers*, afin de pouvoir fédérer des services ayant les mêmes caractéristiques et construire des solutions multi-*cloud*.

LES SERVICES FÉDÉRÉS

Le multi-*cloud*

Les entreprises ont besoin de pouvoir choisir les services numériques qui correspondent à leurs besoins. Elles doivent pouvoir comparer les offres et évaluer les risques qu'elles prennent aussi bien d'un point de vue sécurité que portabilité.

Pour une meilleure compétitivité et indépendance technologique, il est nécessaire de disposer d'une gamme de choix parmi plusieurs fournisseurs de *cloud*, mais aussi de règles claires et vérifiables telles que l'utilisation de données pour les entraînements. Pour le *cloud*, l'Europe dispose déjà d'acteurs de référence. Mais le plus grand d'entre eux ne dépasse pas 2 % de parts de marché au niveau mondial. Les stratégies multi-*cloud* des entreprises ne prennent généralement pas en compte la composante juridique et s'appuient souvent sur des fournisseurs de la même origine géographique comme des *hyperscalers* américains par exemple. Ils ne s'affranchissent pas de possibles pressions politiques, (par exemple, fermetures de services en Russie après l'invasion en Ukraine par les sociétés américaines) ou de risque de perte d'autonomie. Il est nécessaire de diversifier les fournisseurs dans les approches multi-*cloud* en incluant des fournisseurs européens à hauteur de 30 ou 50 %. Cela permet également de répondre à certains besoins réglementaires (capacité d'opérer des banques systémiques ou d'autres infrastructures critiques par exemple).

Pour les données, des échanges de données existent déjà pour de nombreux cas d'usages. Ils sont généralement couverts par des contrats concernant les acteurs de ces échanges. Les identités des partenaires sont reconnues, leur nombre étant restreint. La confiance est possible du fait du peu d'acteurs, et de la gouvernance fermée autour de ces cas d'usages. Par exemple, l'automobile peut avoir créé un espace de données, sous des règles de gouvernance connues de tous ceux qui ont accepté d'en faire partie. Cela peut aussi être le cas d'un autre espace de données existant dans le domaine de l'énergie. Ils fonctionnent en silos et ne sont pas interopérables s'ils ne sont pas construits sur des bases communes.

La Figure 2 (*cf.* page suivante) montre les évolutions et les besoins soulevés par des espaces de données interopérables

L'architecture

Pour répondre aux besoins des espaces de données, l'architecture de Gaia-X repose sur trois éléments fondamentaux.

The demand for Data Spaces

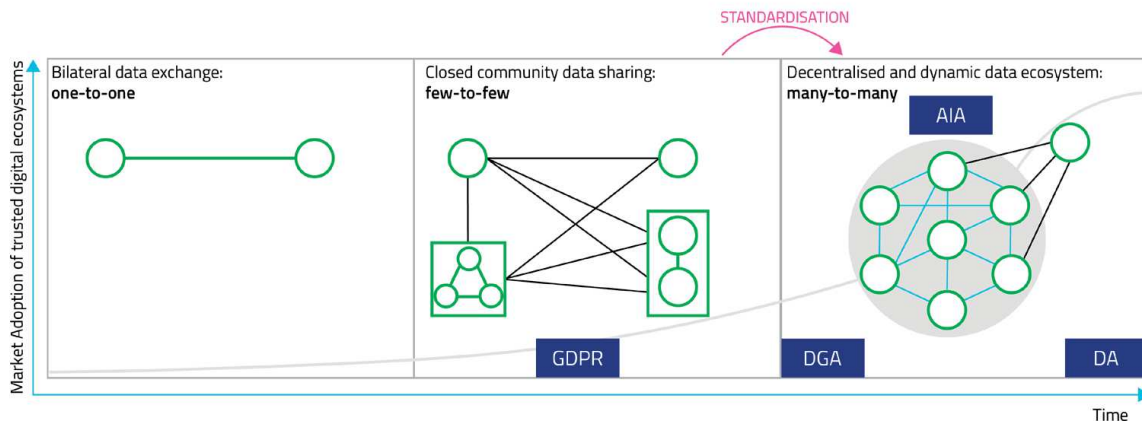


Figure 2 : La décentralisation est essentielle pour le passage à l'échelle (Source : Data Spaces Business Alliance).

Premièrement, le vocabulaire et les définitions de la norme ISO/IEC 17200:2020, aussi appelée CASCO, standardise l'évaluation de la conformité. Cette norme est utilisée pour structurer l'information nécessaire lors de négociation entre deux ou plusieurs parties, notamment les déclarations et les preuves. Cette norme est également la base des certifications existantes dans l'industrie : HDS, PCI-DSS, SecNumCloud, DORA, PSSI, ISO27001, SOC...

Deuxièmement, ces déclarations et preuves sont transformées et encapsulées dans des conteneurs chiffrés, appelés *verifiable credential*. Chaque *credential* peut être vérifié de manière chiffrée, garantissant l'intégrité et la traçabilité des informations. Toute altération non autorisée des informations peut être détectée, même après avoir été partagée plusieurs fois.

L'information contenue dans ces *credentials* est structurée afin de créer un réseau global d'information, un graphe de connaissance, qui utilise les standards et outils des Linked Data, tels que des ontologies et des règles de sémantiques. Cela assure que les déclarations et les preuves sont formulées de manière claire et uniforme, facilitant leur compréhension et utilisation.

Les deux points précédents permettent à chaque partie, utilisateurs, fournisseurs, intermédiaires, d'émettre, d'interroger, de présenter des déclarations ou preuves, quels que soient leurs lieux de stockage, et sans avoir à les dupliquer.

Troisièmement, afin que les déclarations et preuves signées, et sémantiquement structurées puissent être acceptées lors de négociations, il faut que les signatures soient légalement pertinentes.

L'ensemble des signataires et chaînes de certificats légitimes pour signer des déclarations et preuves, appelés Trust Anchors, sont répertoriés sous la gouvernance d'une autorité, dans un registre, appelé Registry.

Ce registre associe pour chaque signataire et chaque chaîne de certificats, un périmètre précis de ce qui est signable. Par exemple, une certification SecNumCloud ne peut être signée que par un organisme accrédité par l'ANSSI.

Ces trois éléments permettent de créer des écosystèmes où les informations peuvent être échangées de manière sécurisée et fiable, tout en respectant les standards de vérifiabilité et de légitimité. Cette architecture assure une gestion efficace et sécurisée des déclara-

tions et des preuves, et permet la création de chaînes traçables d'informations vérifiables en cas de futurs litiges.

Note : La signature électronique est un terme général désignant tout processus électronique permettant de signer un document, tel que l'inclusion de l'image d'une signature manuscrite, tandis que la signature numérique est une forme spécifique de signature électronique qui utilise des techniques de chiffrement pour garantir l'authenticité et l'intégrité du document signé.

Extension par domaine

La section précédente décrit une architecture basée sur trois éléments :

- des schémas de conformité pour apporter des déclarations et des preuves ;
- des déclarations et des preuves organisées suivant des ontologies et modèles sémantiques, puis chiffrées ;
- un registre de quelle entité ou de quel matériel de chiffrement a l'autorité pour signer telles ou telles déclarations ou preuves.

Cette architecture permet d'adresser les couches d'interopérabilité organisationnelle et sémantique, telles que décrites dans l'European Interoperability Framework³.

Gaia-X définit également un ensemble de protocoles et formats de données pour permettre une implémentation technique afin d'opérationnaliser la conformité Gaia-X ou *Gaia-X Compliance* en anglais, décrite dans la section sur les labels.

Cette même architecture peut être utilisée pour soit étendre la conformité Gaia-X à un domaine qui partage les mêmes principes, soit créer de nouvelles règles pour un domaine spécifique ou adaptées à des juridictions extra-européennes.

LA CIRCULATION DES DONNÉES

Les espaces de données sont des acteurs d'un même domaine, travaillant sur des cas d'usages et qui décident de partager de la donnée. Ils portent une gouvernance décidée par ces mêmes acteurs sur les conditions d'accès à ses services. Ils permettent une circulation de données, c'est-à-dire des échanges de données où les propriétaires de ces données vont pouvoir décrire, valider et vérifier les usages qui peuvent être fait de leurs données. Ce n'est que sur ces bases de confiance que cette circulation aura lieu.

Introduction

Comme présenté en introduction, les entreprises doivent de plus en plus travailler en filière pour rester compétitives. Et si elles ne veulent pas être bouleversées par des services numériques développés par les géants du numériques, elles doivent travailler en écosystème pour pouvoir développer des cas d'usages innovants.

Le tourisme, par exemple, a connu un changement majeur de comportement sur les réservations d'hôtels. La mobilité aussi cherche à développer des services qui captent la totalité

³ https://ec.europa.eu/isa2/eif_en/

d'un trajet. Pour recharger un véhicule électrique, il est intéressant de pouvoir affirmer que l'électricité a été produite par des énergies non carbonées. Bref, un ensemble de cas d'usages nécessite la circulation de données. Pour que cette circulation de données existe, le producteur de données doit pouvoir avoir les garanties de confiance qu'il souhaite et pouvoir contrôler les autorisations.

Il est aussi indispensable d'être interopérable entre domaines. Typiquement, un espace de données travaillant dans la mobilité doit pouvoir interagir avec un autre espace de données qui travaille sur l'énergie.

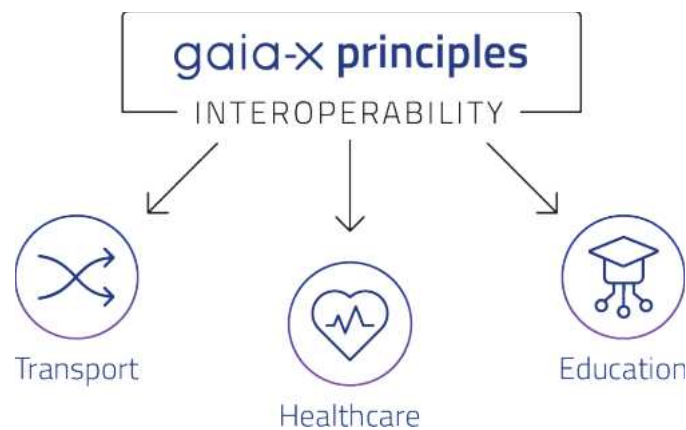


Figure 3 : Interopérabilité entre les différents domaines (Source : Gaia-X).

L'impact de la réglementation

Dans un contexte de mondialisation et de création d'espaces de données impliquant des acteurs non-européens, Gaia-X définit quatre niveaux de conformité : le premier applicable indépendamment de toute législation, les trois suivants appelés Gaia-X Label, et décrits dans la section suivante.

Concernant le premier niveau de conformité, Gaia-X garde intactes les valeurs fondamentales de l'Europe en conservant l'intention de la législation sans nécessairement toujours y faire référence.

Prenons par exemple les deux critères suivants, dont les objectifs sont similaires :

- « Indépendamment de son emplacement et de l'emplacement des utilisateurs du service, un fournisseur de services doit se conformer au RGPD. »
- « Si des données personnelles ou sensibles sont traitées par un prestataire de services, ce dernier doit toujours être en mesure de prouver, sur demande, que le propriétaire des données a donné son consentement sans équivoque pour leur traitement et pour des finalités et une durée explicite. »

La première formulation crée un verrou réglementaire tout en n'apportant que peu de valeur ajoutée car le RGPD est un règlement européen déjà obligatoire et Gaia-X n'est pas un organisme de contrôle législatif. De plus, il existe de nombreux pays où des réglementations similaires en matière de protection des données sont déjà en place : Japon, Brésil, Singapour, États-Unis/Californie, États-Unis/Virginie, etc. Enfin, la formulation ne précise aucunement quelles preuves doivent être fournies.

La seconde formulation détaille davantage la sémantique et la syntaxe des informations à collecter pour démontrer le respect du critère, telles que le consentement, les finalités du traitement, la durée, le point de contact du fournisseur de services pour les demandes d'information, la révocation du consentement et ainsi de suite, afin de parvenir à l'interopérabilité sémantique entre les réglementations existantes en matière de protection des données.

Cet effort supplémentaire de formulation a été fait sur tous les critères du premier niveau de conformité Gaia-X, pour détailler autant que possible la sémantique et la syntaxe des informations requises pour fournir la preuve qu'une exigence est remplie.

Les labels

La confiance est l'élément clé de la création de ces écosystèmes d'infrastructures et de données. Elle doit être objectivée par des critères qui soient mesurables et/ou auditables.

L'association Gaia-X a publié un document de compliance⁴.

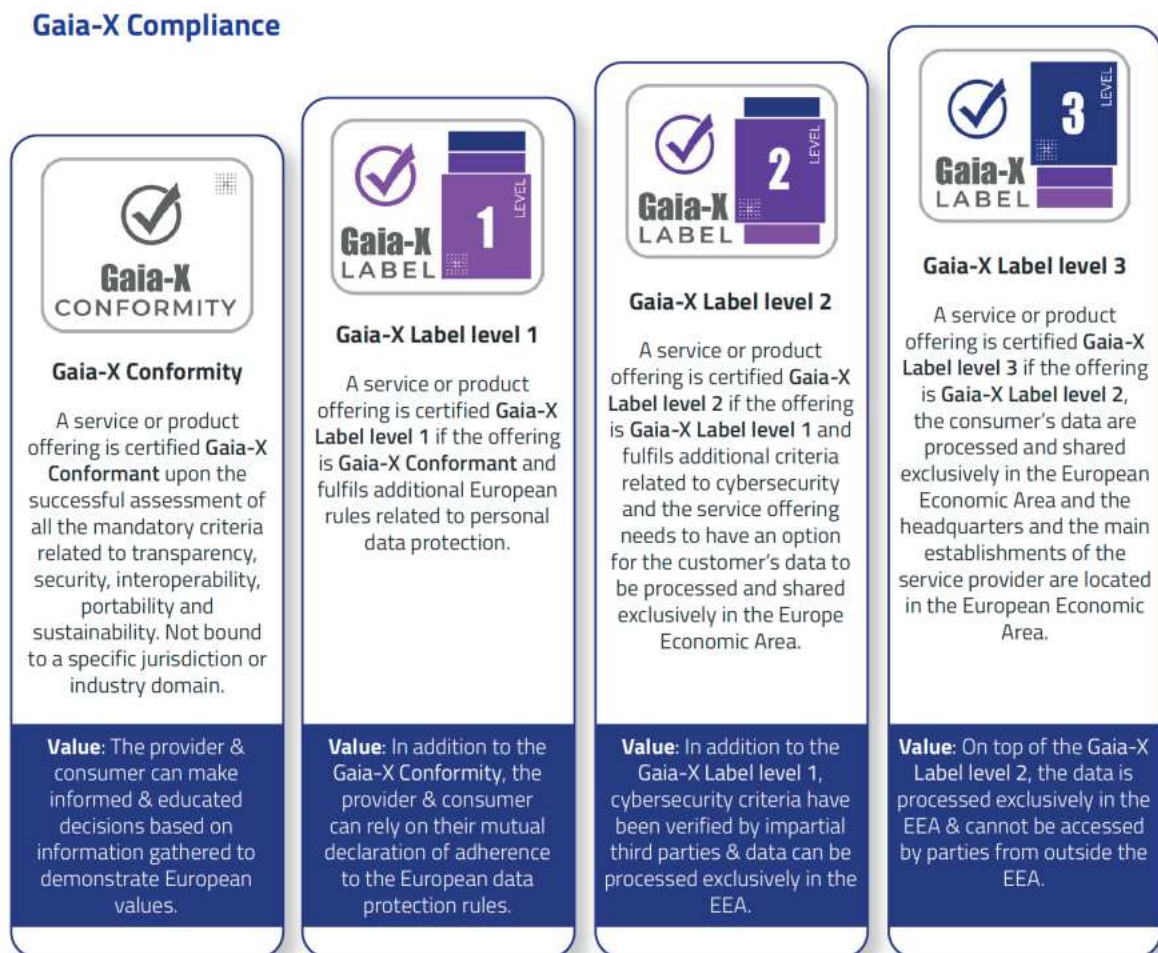


Figure 4 : Les niveaux de confiance de Gaia-X (Source : Gaia-X).

⁴ <https://docs.gaia-x.eu/policy-rules-committee/compliance-document/latest/>

La conformité à Gaia-X et les labels sont applicables aux services des *cloud services providers*. Un travail est en cours dans l'association pour les étendre à des services d'intermédiaires et de partage de données.

Afin qu'un service soit conforme aux principes de Gaia-X et donc de confiance, il doit répondre à un certain nombre de critères. Ce service pourra alors être reconnaissable et sa confiance pourra être évaluée par celui qui souhaite y accéder. L'utilisateur peut alors comparer les offres de services sur des critères communs et choisir celui qui correspond à ses besoins.

Les principes de base des labels de Gaia-X :

- ils sont optionnels, c'est à dire qu'il n'est pas obligatoire d'afficher un niveau de label pour faire partie de l'écosystème Gaia-X ;
- ils sont européens, puisque liés à des réglementations ou valeurs européennes ;
- plus ils sont élevés, plus ils doivent être vérifiés par des *conformity assessment bodies* (CAB) qui vont auditer impartialement les services.

Il existe trois niveaux de labels, comme décrits dans la Figure 4 ci-dessus :

- Le premier niveau est déclaratif, il correspond au niveau de base de conformité auquel s'ajoute l'engagement de respecter les réglementations européennes.
- Le deuxième niveau couvre le niveau 1 et vient ajouter des critères de cybersécurité plus exigeants et validés par un CAB. Les services de niveau 2 doivent aussi obligatoirement proposer une option où les données sont stockées et calculées en Europe.

Le troisième niveau doit être entièrement validé par un CAB et est encore plus exigeant que le niveau 2 d'un point de vue cybersécurité. Il apporte aussi le plus haut niveau d'immunité à des lois extraterritoriales non européennes.

Ce document a été réalisé par les membres et validé par le conseil d'administration de Gaia-X. Il contient des critères qui sont actuellement discutés au sein de EUCS (schéma européen de certification en cybersécurité). L'implémentation des labels est prévue pour le dernier trimestre 2024.

CONCLUSION

L'application à l'AI

L'Europe a créé un ensemble de législations autour des services numériques des plateformes, des données personnelles, de la gouvernance des données des entreprises et maintenant autour de l'intelligence artificielle.

L'IA générative a bouleversé la donne et touché toute la population européenne. Il a été important pour l'Europe de montrer assez rapidement qu'elle allait encadrer les services issus de ces algorithmes.

Il est évident que les données, ayant servi aux apprentissages des IA, sont importantes puisqu'elles peuvent apporter de la précision ou du biais. Les fournisseurs de ces données (les entreprises) doivent pouvoir donner leurs autorisations et vérifier qu'elles sont respectées.

La traçabilité des données et la vérification des autorisations constituent le cœur des principes de Gaia-X.

Le Hub France de Gaia-X a publié un *position paper*⁵ à ce sujet en juin 2024 afin de mettre en lumière la valeur de Gaia-X alors que l'AI Act est mis en application.

Gaia-X à 10 ans

Dans une perspective à 10 ans, Gaia-X pourra se conformer aux futures réglementations, notamment européennes telles que l'AI Act, assurant ainsi une conformité avec les besoins du marché.

Elle offrira une couverture complète, englobant à la fois les fournisseurs et les consommateurs de données, garantissant une gestion sécurisée des informations. De plus, la solution intégrera les besoins des fournisseurs et consommateurs de services, assurant une transparence et une confiance mutuelle dans l'écosystème numérique.

⁵ <https://www.gaia-x-hub.fr/nouveau-position-paper-du-hub-france-de-gaia-x/>

Les réseaux virtualisés : promesses et enjeux

Par Fabrice GUILLEMIN
Orange

Les technologies de virtualisation ont révolutionné ces dernières années les réseaux de télécommunications. Au-delà des avancées techniques et de la très grande flexibilité apportée, cette rupture dans la conception des réseaux pose de nombreux défis, en particulier vis-à-vis des infrastructures sous-jacentes, le *cloud* notamment, et l'exploitation des réseaux. Cet article vise à faire le point sur des sujets sensibles pour les opérateurs de réseaux et à identifier des solutions possibles.

VIRTUALISATION DES RÉSEAUX

Les technologies de virtualisation des systèmes informatiques ont révolutionné l'architecture des réseaux de télécommunications en dissociant le *software* du *hardware* d'hébergement. Cette transformation permet de rendre les réseaux programmables, avec des versions logicielles mises à jour en continu. Les fonctions réseau sont particulièrement adaptées à cette approche, donnant naissance au concept de virtualisation des fonctions réseau (Network Function Virtualisation, NFV), initialement développé par l'ETSI [1]. L'application de NFV à l'ensemble des fonctions réseau (contrôle, transfert, routage, commutation) a mené à l'émergence des réseaux définis par le logiciel. Le concept de SDN (*Software Defined Networks*) se concentre sur la programmation du plan de transfert *via* le protocole *OpenFlow* [2].

L'approche NFV s'applique à des fonctions aussi variées que les *firewalls* ou les cœurs de réseau mobile, qui comportent eux-mêmes de nombreux composants logiciels interagissant entre eux, comme le cœur de réseau 5G. Pour augmenter la flexibilité de l'approche NFV, il est possible de décomposer les fonctions ou leurs composants en des entités plus petites, chacune réalisant une tâche élémentaire. Cette méthode, empruntée au monde informatique et correspondant à l'approche dite par microservices, est particulièrement pertinente pour les réseaux.

Dans ce contexte de NFV, les microservices sont généralement hébergés par des conteneurs, dans des infrastructures *cloud* gérées par des outils spécialisés. Kubernetes (K8S) [3] est couramment utilisé pour gérer des conteneurs, bien que des méthodes manuelles (comme Docker Compose) soient aussi possibles. L'avantage de Kubernetes réside dans son ensemble d'outils facilitant l'automatisation du déploiement de conteneurs sur des *clusters* de serveurs. Il permet en outre le redimensionnement automatique des conteneurs selon l'activité des microservices (*autoscaling*).

La combinaison de la virtualisation des fonctions réseau et de la décomposition en microservices hébergés dans des conteneurs gérés par Kubernetes représente l'approche *cloud* native, qui semble actuellement la mieux adaptée aux réseaux. Développée notamment par Google, cette approche permet de déployer et modifier les microservices en continu grâce à une chaîne d'intégration et de développement continu (CI/CD). Les projets logiciels développant les fonctions réseau archivent majoritairement le code sur des réper-

toires Git et ceux-ci peuvent être utilisés pour déployer automatiquement les fonctions *via* GitOps [4]. Ce dernier forme un cadre opérationnel basé sur les bonnes pratiques de DevOps en étendant l'approche *Infrastructure/Network as Code* (I/NaC), qui automatise le déploiement des infrastructures IT à partir de fichiers de configuration.

Ces différentes approches visent à rendre les réseaux plus flexibles et dynamiques, automatisant le déploiement et la mise à jour des fonctions réseau virtualisées, tout en permettant une gestion autonome des microservices et la reconfiguration des réseaux. Toutefois, elles bouleversent les pratiques traditionnelles des opérateurs, qui privilégient des mises à jour par paliers fonctionnels, une validation avant déploiement et une gestion majoritairement manuelle. Les constructeurs traditionnels, qui développent des fonctions réseau optimisées sur du *hardware* spécifique, doivent également s'adapter pour rendre les fonctions agnostiques au *hardware*. Cela soulève des défis opérationnels, de compétences et de conception, nécessitant une attention particulière à l'infrastructure *cloud* qui supporte ces nouvelles technologies.

INFRASTRUCTURES CLOUD

Dans la transformation des réseaux en systèmes entièrement logiciels et dynamiques, l'infrastructure *cloud* joue un rôle crucial. Au début de NFV, les machines virtuelles (VM) étaient privilégiées, mais elles se sont rapidement révélées moins performantes que la technologie de conteneurisation, plus légère et plus facile à déployer. Bien que l'hébergement de *clusters* K8S dans des VM soit possible et offre une isolation supérieure comparée à une implémentation directe sur le système d'exploitation d'un serveur (solution dite *bare metal*), il est maintenant communément admis que l'infrastructure *cloud* la plus adaptée pour les réseaux logiciels devrait reposer sur des *clusters* K8S.

Pour les fonctions de contrôle qui n'ont pas d'exigences de temps réel strictes, comme les fonctions de contrôle (authentification, bases de données clients, etc.), les *clouds* centralisés suffisent. Il est courant de voir des déploiements de réseaux mobiles privés utilisant les *clouds* d'Amazon, de Microsoft ou de Google. Cependant, certaines fonctions du plan de transfert des données avec des contraintes fortes en termes de débit et de latence doivent être implantées au plus près des utilisateurs, en particulier la passerelle vers Internet. En raison du manque de standardisation entre les plans de contrôle et de données, certaines fonctions de ces derniers sont souvent colocalisées dans des implémentations commerciales ou *open source*, telles que Magma.

Outre les fonctions du cœur de réseau, celles du réseau d'accès sont également candidates à la virtualisation. L'Alliance Open RAN (ORAN) travaille à la désagrégation des unités de traitement radio (BBU) en différentes fonctions : CU (*Centralized Unit*), DU (*Distributed Unit*) et RU (*Remote Unit*) [6]. Les fonctions de contrôle radio (CU) sont clairement éligibles à la virtualisation, tandis que les DU, qui traitent le codage et le *scheduling* radio, nécessitent des capacités de calcul plus importantes. Les RU sont encore principalement implémentées sur des FPGA¹ dédiés. La virtualisation des fonctions CU et DU, regroupées dans des *clouds* communs pour des raisons énergétiques et de gestion de trafic, s'inscrit dans l'approche *cloudRAN*.

Cette tendance à virtualiser certaines fonctions du plan de transfert conduit au déploiement d'infrastructures *cloud* en bordure de réseau, appelées *edge cloud*. Ce terme est interprété différemment selon les acteurs : les GAFAM l'utilisent pour désigner des *data centers* régionaux à leur échelle mondiale, tandis que les opérateurs nationaux l'associent à des *data centers* régionaux au sein d'un même pays. Pour ces derniers, la

¹ *Field-programmable gate array* : matrice de portes programmables. Voir par exemple : <https://www.futura-sciences.com/tech/definitions/technologie-fpga-8700/>

virtualisation du réseau d'accès et de certaines fonctions du cœur de réseau nécessite le déploiement de *data centers* à travers tout un pays, notamment pour les fonctions CU et UPF, voire plus proches des antennes (à une centaine de kilomètres) pour les DU, qui peuvent nécessiter des accélérations *hardware* ou des GPUs.

La virtualisation des fonctions réseau représente donc un investissement considérable pour les opérateurs, déjà soumis à la pression des GAFAM qui maîtrisent les technologies du *cloud* et ont acquis des *start-ups* spécialisées dans les fonctions réseau. Investir massivement dans une infrastructure *cloud* distribuée pose un défi stratégique majeur pour les opérateurs. Pour rentabiliser cet investissement, ils doivent trouver des applications nécessitant un déploiement en bordure de réseau et générant des revenus substantiels. Cependant, la latence d'un réseau national, qui est de l'ordre de quelques dizaines de millisecondes, est compatible avec la plupart des applications actuelles, y compris les jeux et l'intelligence artificielle. Une politique de *peering* adéquate avec les acteurs OTT pourrait être une solution économique pour offrir des applications sensibles à la latence. Ces acteurs supporteraient les coûts d'investissement, mais bénéficieraient des revenus directs ou indirects des applications.

Cette situation rend la problématique du *edge cloud* complexe pour les opérateurs de réseau nationaux. Les causes de latence pour les applications sont souvent localisées dans les zones radio (cellulaires et wifi), et le déploiement de *data centers* très proches des utilisateurs n'améliore la latence que marginalement. Cependant, le traitement de grandes quantités de données, comme les flux vidéo de télésurveillance nécessitant des débits d'*upload* élevés, peut représenter un cas d'usage intéressant pour le *edge cloud*. Le problème est de trouver une équation économique équilibrée pour les opérateurs.

Pour conclure, la virtualisation des fonctions réseau soulève des préoccupations en matière de sécurité, notamment vis-à-vis des données sensibles des utilisateurs. L'hébergement de ces fonctions dans des *clouds* publics comme GCP ou AWS pose des problèmes de confidentialité des données (RGPD en Europe). Si ce problème peut être non critique dans un environnement privé, comme un réseau privé d'une multinationale, il devient beaucoup plus préoccupant dans le contexte des réseaux de télécommunications publics. Les recommandations actuelles de l'ANSSI n'autorisent pas ce genre de déploiement. Pour surmonter ce problème, il est nécessaire de développer une solution de *cloud* maîtrisé par l'opérateur de réseau (cf. le projet Sylva [7]).

ORCHESTRATION

Un défi connexe à la virtualisation des réseaux et à la capacité de créer un grand nombre de réseaux virtualisés sur un même substrat est la manière d'orchestrer ces déploiements. L'orchestration se distingue de l'automatisation, cette dernière visant à automatiser des tâches auparavant manuelles, tandis que l'orchestration coordonne et organise différents déploiements pour atteindre des objectifs globaux d'optimisation pour l'infrastructure (par exemple, taux d'occupation des serveurs, utilisation des liens de transmission) et pour les réseaux virtuels déployés (par exemple, latence, taux de perte de paquets).

Les outils d'orchestration intègrent des bibliothèques de fonctions à déployer, des interfaces de commande (interface « nord »), des inventaires, des modules de collecte de données et d'optimisation, et parfois des méthodes pour définir des règles métier permettant de déployer des boucles autonomes de contrôle. La plateforme ONAP (Open Network Automation Platform) a été conçue dans cette optique, offrant une vue holistique sur tous les segments du réseau (réseaux d'accès, de transport, de cœur et l'infrastructure *cloud*) [8]. ONAP comprend tous les modules nécessaires pour optimiser à la fois le substrat (infrastructures *cloud*, réseaux de transport et de routage sous-jacents) et les réseaux virtuels déployés. Les techniques d'intelligence artificielle se révèlent en particulier très efficaces pour résoudre les problèmes d'optimisation associés [9].

Cependant, concevoir une plateforme unique globale est difficile et conduit à des solutions lourdes à déployer et à maintenir. De plus, chaque composante du réseau tend à développer ses propres plateformes d'orchestration, en particulier les plateformes *cloud* avec K8S et l'automatisation *via* GitOps. Étant donné que les réseaux sont naturellement répartis géographiquement, par exemple avec des réseaux d'accès (RAN) et cœurs virtualisés, il est nécessaire d'avoir des versions de K8S *multi-cluster* pour déployer un réseau virtualisé de manière cohérente. Plusieurs projets, comme Nephio, travaillent sur des versions de K8S *multi-cluster*.

En dehors des aspects relevant du *cloud*, différents segments du réseau développent également leurs propres plateformes d'orchestration. Par exemple, les RIC (*RAN Intelligent Controller*) d'ORAN (temps réel et non temps réel) visent à orchestrer et optimiser les réseaux d'accès radio. Les fonctions réseau virtualisées, comme les cœurs de réseau, sont également conçues avec leurs propres fonctions d'orchestration, même si celles-ci sont le plus souvent des plateformes d'administration propres à la fonction. L'orchestration de bout en bout se limite alors à coordonner les différents orchestrateurs pour déployer des réseaux virtuels de bout en bout, sachant que certaines tâches nécessitent de toute façon des tâches manuelles, ne serait-ce que l'installation d'antennes radio pour déployer un réseau radio. L'architecture *Open Digital Architecture* (ODA) du TM Forum [10] modélise les différentes parties d'un réseau et les tâches d'orchestration associées. Néanmoins, une orchestration cohérente de bout en bout reste encore aujourd'hui un défi majeur pour les opérateurs. Une multitude de sociétés apparaissent pour assurer l'intégration du code et opérer des outils pour orchestrer les réseaux de bout en bout.

CONCLUSION

Au-delà des promesses apportées par les technologies de virtualisation pour les réseaux de télécommunications, celles-ci bouleversent totalement l'écosystème et le métier des opérateurs de réseau s'en trouve totalement modifié. Après la vague d'enthousiasme soulevée à leur émergence rapide, la maîtrise opérationnelle de ces technologies reste en grande partie à consolider, surtout pour les opérateurs de réseau historiques. L'écosystème industriel est lui-même en continuelle transformation pour essayer de trouver un point d'équilibre sur ces technologies.

Sur le marché grand public, les opérateurs classiques restent les mieux placés par leur proximité avec les clients ; la virtualisation et toutes les techniques du *cloud* peuvent leur apporter beaucoup de flexibilité même si la maîtrise globale reste un défi majeur, surtout vis-à-vis de l'intégration de très gros logiciels. Sur le marché des réseaux privés cependant, la tension avec les GAFAM risque de s'accroître étant donné qu'ils se sont donné les moyens d'offrir des solutions performantes aux clients professionnels. Les opérateurs de réseau peuvent néanmoins utiliser la proximité et de l'assistance qu'ils apportent à leurs clients pour rester dans la chaîne de valeur.

RÉFÉRENCES

- [1] YI B., XINGWEI W., LI K., DAS S. K. & HUANG M. (2018), "A comprehensive survey of Network Function Virtualization", *Computer Networks*, vol. 133, pp. 212-262.
- [2] XIA W., WEN Y., FOH C. H., NIYATO D. & XIE H. (2015), "A Survey on Software-Defined Networking", in *IEEE Communications Surveys & Tutorials*, vol. 17, n°1, pp. 27-51.
- [3] BURNS B., GRANT B., OPPENHEIMER D., BREWER E. & WILKES J. (2016), "Borg, Omega, and Kubernetes", *ACM Queue*, vol. 14, pp. 70-93.
- [4] BEETZ F. & HARRER S. (2022), "GitOps: The evolution of DevOps?", *IEEE Software*, 39(4), IEEE Computer Society Press, <https://doi.org/10.1109/MS.2021.3119106>

[5] RIVERA D., MORENO J., SANZ RODRIGO M., LOPEZ D. & MOZO A. (2023), “Providing heterogeneous signaling and user traffic for 5G core network functional testing”, *IEEE Access*, vol. 11, pp. 2968-2980.

[6] POLESE M., BONATI L., D’ORO S., BASAGNI S. & MELODIA T. (2023), “Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges”, *IEEE Communications Surveys and Tutorials*, 25(2).

[7] <https://sylvaproject.org/>

[8] SLIM F., GUILLEMIN F., GRAVEY A. & HADJADJ-AOUL Y. (2017), “Towards a dynamic adaptive placement of virtual network functions under ONAP”, 2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Berlin, Germany, pp. 210-215.

[9] ALVES ESTEVES J. J., BOUBENDIR A., GUILLEMIN F. & SENS P. (2022), “A heuristically assisted deep reinforcement learning approach for network slice placement” in *IEEE Transactions on Network and Service Management*, vol. 19, n°4, pp. 4794-4806.

[10] <https://www.tmforum.org/oda/>

La *softwarization* des réseaux et son impact technico-économique

Par Emmanuel PUIG
Amarisoft

La *softwarization*, ou migration des fonctions informatiques du matériel vers le logiciel, marque une évolution technologique clé où les CPU de plus en plus puissants permettent des solutions *software* autrefois impossibles. Cette transition, visible dans des domaines variés comme les jeux vidéo et les télécoms, est motivée par des avantages économiques et opérationnels tels que la flexibilité, la réduction des coûts, et l'accès à une gamme étendue d'outils et de ressources *open source*. Dans les réseaux télécoms, bien que les cœurs de réseau soient déjà largement « *softwarisés* », l'accès radio reste majoritairement *hardware*, mais la tendance change avec l'essor des réseaux privés et des technologies comme la 5G, favorisant des solutions plus flexibles et adaptées aux petits volumes. Les entreprises qui adoptent cette transformation au moment opportun peuvent gagner un avantage concurrentiel significatif.

LA *SOFTWARIZATION*, UNE PREUVE DE MATURITÉ

Le débat entre le *software* et le *hardware* est un débat presque aussi ancien que l'informatique. Si la distinction entre le logiciel et le matériel informatique est discutable et que sa frontière est parfois floue, nous nous placerons dans le cadre où une implémentation *software* sera du code destiné à être exécuté sur un CPU (Processeur à Usage Général) et une implémentation *hardware* sera sur du silicium dédié à la tâche.

Ce choix a toujours été une affaire de compromis, donc soumis aux conditions technologiques à un moment donné. Cela implique par conséquent que ce choix peut être remis en cause avec l'évolution de ces conditions.

Lorsque l'on s'intéresse à l'évolution de ces conditions, on remarque que pour une nouvelle technologie, le premier choix est bien souvent le choix *hardware* pour la simple raison que la puissance de calcul disponible sur les CPUs est largement insuffisante et qu'il est donc fonctionnellement impossible de faire avec.

La loi de Moore aidant, il arrive un moment où la capacité des CPUs rend possible une implémentation *software*.

À ce stade, le facteur économique penche en général largement vers la solution *hardware* : les machines à base de CPUs capables de fournir le même service étant en général beaucoup plus chères.

Avec un peu plus de temps, ce frein finit par ne plus en être un et d'autres critères peuvent alors rentrer en jeu. Ainsi, les avantages du développement *software* finissent par l'emporter sur ceux du *hardware*. La contamination se fait par l'entrée de gamme et monte au fur et à mesure que les CPU rattrapent leur retard.

Les exemples de ce phénomène de fond ne manquent pas : dans le jeu vidéo, on est passé des bornes d'arcades, assemblage de puces dédiées aux consoles d'aujourd'hui qui ne sont plus finalement que des PCs. De la même manière, le monde de l'audio vidéo est passé d'encodeur et de décodeurs dédiés à des suites logicielles représentées par les très français FFmpeg et VLC.

Dans le domaine des télécoms, l'exemple le plus marquant a été l'avènement du logiciel Asterisk reléguant de gros et dispendieux équipements au passé.

Bien évidemment, pour qu'une telle transition ait lieu, il est indispensable que les solutions *software* soient fonctionnelles (sinon aucun débat n'est possible) et aient un niveau de maturité suffisant, auquel cas l'adoption d'une approche *software* devient une vraie question.

On notera que cette description doit bien évidemment être soumise aux subtilités de chaque technologie et qu'en pratique, il n'y a pas de critère absolu.

LES AVANTAGES DE L'APPROCHE *SOFTWARE*

Pourquoi vouloir adopter une approche *software* plutôt qu'*hardware* ?

Il y a d'abord les promesses, réelles et parfaitement connues et documentées que sont les possibilités de virtualiser, d'automatiser simplement, de désagréger, de mutualiser dans des *data centers*, d'offrir de nouveaux *business models*...

Ici, nous resterons pragmatiques et opérationnels en nous focalisant sur l'aspect coût du cycle de vie d'un produit, et en raisonnant toutes choses égales par ailleurs, sans avoir à mettre dans la balance des telles promesses, nécessitant des modifications importantes et débordant l'aspect technique.

Dans le développement d'un nouveau produit, on distinguera : les coûts de développement initiaux, principalement de R&D, les coûts de production du produit (usine, logistique...), ceux de son exploitation (maintenance, opérations) et ceux liés à son évolution (nouvelles fonctionnalités, améliorations).

Toutes ces phases vont bénéficier de la *softwarization*, en les accélérant, les simplifiant, les rendant moins dépendantes...

Comme dans tout processus d'ingénierie, dans une démarche empirique, il faut itérer un certain nombre de fois. Chaque itération sera coûteuse en temps et en ressources. Par la nature même du *software*, une itération peut être plus rapide de plusieurs ordres de grandeur : une compilation sera toujours plus rapide que produire un prototype. À l'heure de l'agilité, cela est fortement appréciable.

Dans un monde où il y a une tension sur la disponibilité d'ingénieurs qualifiés, les ressources disponibles seront plus accessibles en *software*, car plus génériques et même s'il sera toujours nécessaire d'avoir une connaissance du domaine, une expertise *software* sera plus facilement adaptable.

Travailler sur des CPUs donne accès à tous les outils possibles et imaginables disponibles en *open source* et permettant d'accroître la productivité et la qualité des produits. Parlez-en à vos équipes, vous pourriez voir des étoiles dans leurs yeux.

Utiliser des composants standards donne accès à une gamme incroyablement étendue de fournisseurs, permettant de réduire les coûts, la dépendance et les possibilités. On appréciera la possibilité de fonctionner en temps non réel en n'étant plus tributaire d'une horloge physique par exemple.

Les mises à jour *hardware* étant coûteuses (rétrofit) contrairement au *software*, plus cette dernière sera prépondérante, plus les possibilités seront importantes. On appréciera par exemple la possibilité de répondre aux problèmes de sécurité.

Cette liste n'est évidemment pas exhaustive, et le calcul économique variera en fonction des situations, et si les avantages ne sont pas suffisants, on restera sur une approche *hardware*, le propos n'étant pas de dire que le *software* est meilleur dans l'absolu. Des notions de performance ou énergétique pouvant être dans certains cas rédhibitoires.

Cela étant dit, il faut bien comprendre qu'au-delà de toutes ces considérations, penser en termes de *software* impose un changement de paradigme :

- dans l'approche *hardware*, les spécifications du produit doivent être définies *a priori* et fixeront un cadre indépassable. Une fois celui-ci planté, les fonctionnalités devront rentrer dans ce cadre ;
- lorsque l'on adopte une approche entièrement *software*, ce cadre saute dans la mesure où le matériel acquiert une forme de flexibilité. Celui-ci peut être modifié *a posteriori*, décliné en différentes versions, mutualisé... Cela est particulièrement prégnant dès que l'on veut bénéficier des économies d'échelle ou pouvoir décliner simplement un produit en gamme.

En définitive, tôt ou tard, l'équilibre finira par pencher en faveur du *software*. La question est donc quand ?

LA MIGRATION, UN PROCESSUS LENT MAIS INEXORABLE

La migration du *hardware* vers le *software* ne se fait que très rarement en une seule fois mais découle plutôt du glissement progressif des différentes fonctions d'un produit.

Nous sommes très loin de l'époque où le *hardware* était entièrement dédié à la tâche à réaliser. Aujourd'hui, la grande majorité des produits repose sur des SoC¹ implémentant des CPUs. Ainsi, de plus en plus de fonctions terminent sur le CPU et ne sont plus réalisées par des macros *hardware*.

Lentement, la balance entre les CPUs et le *hardware* spécialisé se déséquilibre en faveur du premier.

Et dans certains cas, tout peut être fait sur le CPU.

SOFTWAREZATION DES RÉSEAUX TÉLÉCOMS

Ce phénomène de fond est à l'œuvre dans le domaine des réseaux filaires depuis plusieurs décennies. Le transport de l'information par paquets a remplacé la commutation de circuit, l'écrasante majorité des fonctions sont l'œuvre de logiciels tournant sur des serveurs Linux dans des data center, et même les composants les plus critiques en termes de performance, comme les *switchs*, exécutent une grande partie de leurs fonctions sur des CPUs. Le *software* a fini par totalement envahir ce secteur et relègue petit à petit le *hardware* à des cas bien spécifiques² [1].

¹ *System on a chip* : il s'agit de puces intégrées, dont les versions récentes embarquent le CPU, le GPU, la puce dédiée au réseau, au wifi et au bluetooth, la RAM et la mémoire de stockage, sur une seule et même puce.

² <https://theses.hal.science/tel-03752344/>

QUEL AVENIR POUR LA *SOFTWARE*IZATION DES RÉSEAUX TÉLÉCOMS ?

Comme expliqué précédemment, le choix du *software* est circonstanciel, il impose donc par essence, afin d'être au plus près de l'innovation, d'être capable de mettre à jour les éléments ayant abouti à ce choix afin de saisir le moment où la balance s'inverse. Et cela ne dépend pas seulement de facteurs techniques, une grande part du processus décisionnel est dépendante du facteur humain, de ses biais, de ses croyances. Il est donc particulièrement important dans ce contexte de ne pas seulement connaître les choix mais de ne jamais perdre de vue les conditions de ce choix, sous peine de croire en cette fameuse phrase faussement attribuée à Mark Twain : « ils ne savaient pas que c'était impossible, alors ils l'ont fait ».

Étant donné les avantages du *soft*, il est clair que les acteurs qui franchissent le pas au moment opportun, c'est-à-dire au moment où l'inéquation s'inverse, en tirent un avantage certain, car non seulement ils seront présents plus tôt, mais ils bénéficieront de l'accélération inhérente au développement *software*.

Il est donc très important pour toute organisation de fournir une veille et même des expérimentations, appelons cela de la recherche, en permanence. Les équipes techniques doivent être challengées et sorties de leur zone de confort afin d'être prêtes à anticiper le mouvement plutôt que de le suivre.

Dans le cadre des réseaux télécoms, et en particulier mobile, les cœurs de réseau sont déjà fortement « *softwarisé* ». En témoigne les architectures 3GPP qui utilisent de manière intensive la désagrégation en éclatant les différents éléments fonctionnels en autant de machines et de logiciels.

Il reste cependant un bastion encore largement dominé par le *hardware* : l'accès radio.

Si le processus a déjà été initié par des projets comme OpenBTS, il relevait encore du bricolage. Mais aujourd'hui nous voyons poindre des solutions matures et compétitives pour de nombreux cas d'usage.

Par exemple, les stations de base sont aujourd'hui largement spécifiées pour le marché des grands réseaux publics, seuls à même de justifier de lourds investissements de par les volumes mis en jeu.

À ce titre on remarquera l'échec du déploiement des *femtocell* et autre dérivés, coincées par les faibles marges.

Avec l'avènement des réseaux privés, par essence parfaitement hétérogènes, le nombre de cas d'usage et leur spécificité explosent. Un réseau sur une plateforme pétrolière ne présente pas du tout les mêmes caractéristiques et besoins qu'un réseau de diffusion itinérant sur un événement sportif.

Si l'on ajoute à cela l'incroyable flexibilité des technologies telle que la 5G, cette problématique devient vertigineuse.

Afin d'exploiter tout ce potentiel et ainsi permettre l'avènement de nouveaux marchés, ce qui fera la différence est la capacité des acteurs à concevoir des produits adaptés pour des petits volumes. Ainsi, l'approche *hardware*, parce que son développement initial est long et coûteux n'est pas adapté, car on ne peut envisager un *hardware* dédié qui serait optimisé pour tous les cas. À partir du moment où l'approche *software* est utilisée, le changement de *hardware* (le PC) n'est plus un problème.

L'adoption récente [2] par les références de l'industrie de CPU Intel x86 dans leurs produits d'accès radio est un indicateur significatif que ce changement de paradigme est déjà à l'œuvre.

RÉFÉRENCES

[1] DLAMINI, T., ROSSI M. & MUNARETTO D. (2017), “Softwarization of mobile network functions towards agile and energy efficient 5G architectures: A survey”, *Wireless Communications and Mobile Computing*, 8618364, 21.

[2] <https://www.ericsson.com/en/press-releases/2023/6/ericsson-first-cloud-ran-call-with-intel-latest-generation-processor-on-hpe-server> ; <https://www.nokia.com/about-us/news/releases/2024/02/21/nokia-collaborates-with-intel-to-deliver-high-performance-cost-efficient-private-5g-networks/> ; <https://news.samsung.com/global/samsung-expands-collaboration-with-intel-to-advance-vran-innovation>

Nouvelles technologies optiques, quel impact sur les systèmes de réalité augmentée de demain ?

Par Jean-Louis de BOUGRENET DE LA TOCNAYE
Professeur, chef du département d'Optique à l'IMT Atlantique

La réalité augmentée (RA) a souvent progressé à la faveur de sauts technologiques. Ce fut le cas avec la miniaturisation des afficheurs, des pico-projecteurs et des combineurs. Les dernières avancées dans le secteur de la conception/fabrication d'optiques, dites à effet de surface (*e.g.* méta-surface, -lentille et -forme) auront-elles le même impact sur les dispositifs de projection des futurs systèmes RA ? Des avancées ont déjà eu lieu dans ce domaine avec les *freeforms*. De la même façon, le développement de la RA holographique, porté par la montée en puissance dans le domaine *consumer* des lunettes intelligentes, imposera-t-il ce choix à l'ensemble des dispositifs RA ? Autant de questions qui seront brièvement abordées dans cet article, consacré à la couche optique des systèmes RA. Enfin, dans cette course permanente vers toujours plus d'immersivité et du plus près de l'œil, les lentilles de contact seront-elles le futur de la RA hyper-immersive ? Quelles sont les dimensions ultimes que l'on pourra raisonnablement atteindre ? Autant de questions que l'on tentera d'aborder ici.

INTRODUCTION

Le déploiement des dispositifs de réalité augmentée (RA) a souvent été rythmé par les avancées technologiques. Employée dès les années 1950 dans le secteur de la défense [1], la RA a dû attendre les années 1990 pour bénéficier de nouveaux concepts d'affichage. Cette décennie sera propice à l'extension dans divers domaines comme l'aéronautique civile [2] (simulateur de vol, écran sur *Head Mounted Display* –HMD) et l'automobile (affichage tête haute *Head up Display* – HUD). Il faudra patienter encore une quinzaine d'années pour voir des HMD RA immersifs à écrans d'affichage pour *smartphones* [3]. L'écosystème technologique du *smartphone*, sa connectivité sèmeront les germes du second boom de la RA, posant les premières briques pour les intégrateurs présents aujourd'hui sur le marché. Ces technologies d'affichage ont été le véritable catalyseur des générations qui suivront jusqu'à nos jours.

L'émergence de dispositifs d'affichage immersifs proches des yeux (*Near-To-Eye*, NTE) va soulever d'autres questions comme l'étendue des sources, la focalisation statique, la gestion de faibles contrastes et luminosités. Ces défis vont mettre l'accent sur les dispositifs de projection qui concentrent aujourd'hui la plupart des efforts, afin de mieux les adapter aux spécificités du système visuel humain. Si l'on regarde les évolutions actuelles en matière de technologies optiques, on peut penser que l'arrivée à maturité des technologies de méta-surfaces [4], qui diffusent aujourd'hui dans de nombreux secteurs, constitue une nouvelle étape de rupture dans l'histoire des dispositifs RA. J'en veux pour preuve la convergence entre technologies silicium et méta-surfaces illustrée par l'association entre Metalenz et STMicroelectronics [5] pour produire les caméras compactes à grande profondeur de champ de nos *smartphones*, démontrant le potentiel de ces nouvelles technologies.

Ceci ne sous-estime en rien le poids des logiciels qui feront des systèmes RA les prochaines plateformes informatiques, remplaçant ordinateurs portables et tablettes. Bien au contraire, cela renforce les exigences de ces plateformes en matière d'optiques compactes (moteur d'affichage, optique de combinaison, etc.). S'il fallait montrer à quel point les aspects matériels et logiciels sont devenus indissociables, je citerais le rapprochement entre Meta et Essilor-Luxottica [6] sur les lunettes intelligentes. Le sujet est donc vaste et complexe, y compris sous l'angle matériel. Je n'ai pas l'ambition de le traiter en entier. J'insisterai sur les blocs optiques de projection (principaux points de blocage) pour montrer comment l'émergence de nouvelles filières technologiques pourrait conduire à des avancées significatives en RA. Je traiterai les deux options majeures de combineurs, cœur des systèmes RA actuels. De nombreuses autres solutions ont été proposées pour contourner les limitations des systèmes optiques, telles que l'approche holographique rendue possible grâce au déploiement de sources laser intégrées, permettant de contourner certains invariants des systèmes d'imagerie. Je finirai enfin avec quelques solutions exotiques comme les lentilles de contact. Pour plus de détails, le lecteur pourra se reporter au très complet ouvrage de Bernard Kress [7], dont je conseille la lecture.

Avant cela, rappelons quelques exigences à l'origine des défis matériels imposés aux dispositifs RA, comme le besoin d'immersion, lié aux perceptions multisensorielles et le confort (réduction du poids, de la taille, recul du centre de gravité), visuel (repères 3D naturels, large FOV, une bonne résolution angulaire), et social. Pour faciliter la compréhension, je rappellerai les ruptures technologiques en train de révolutionner la conception et la fabrication optique et auxquelles je ferai appel par la suite.

LES OPTIQUES MINCES : *FREEFORM*, META-SURFACE, -LENTILLE, -FORME

De façon simplifiée une surface *freeform* est une surface dont la forme n'a pas de symétrie de translation ou de rotation par rapport aux axes normaux au plan moyen [8]. Cela affranchit la conception optique. Les surfaces traditionnelles étant sphériques ou asphériques à symétrie de rotation, l'usage d'optique anamorphique implique la combinaison avec des surfaces toroidales. Les *freeforms* permettent un traitement naturel des aberrations [9], ce qui est un élément important du confort visuel en RA.

Les méta-surfaces sont de nature très différente. Ce sont des versions planaires minces des métamatériaux [4], constituées de nano-éléments dont la dimension est inférieure à la longueur d'onde. Elles permettent de bloquer, absorber, concentrer, disperser, guider les ondes. Ce sont des effets de surface. Pour simplifier, le contrôle de la taille et de la forme des cellules unitaires (les méta-atomes [4]) permet d'obtenir plusieurs indices effectifs et créer différents comportements de surface. Les méta-lentilles sont une extension des méta-surfaces, appliquée à des profils de phase non uniforme ou périodique, comme c'est le cas pour une lentille (voir la Figure 1a). Cela se fait par une micro-structuration au moyen de groupes de méta-atomes. Ceci permet de concevoir des optiques à très grandes ouvertures numériques ($NA > 0,8$), corriger des aberrations (e.g. chromatiques), concevoir des optiques polarisantes, le tout à partir de surfaces minces [4], avantage clé en RA. Pour plus de détails le lecteur pourra consulter l'ouvrage de référence de H. Benisty, J.-Y. Geffret et Ph. Lalanne [11].

Les méta-formes [10] (Figure 1b) sont des surfaces optiques combinant les avantages d'une *freeform* et d'une méta-surface en un même élément. Elles répondent aux critères de systèmes optiques à haute résolution, avec un facteur de forme compact, tels les écrans de RA comme on le verra. La nouveauté est que ces technologies sont manufacturables de nos jours, y compris dans la bande visible [11].

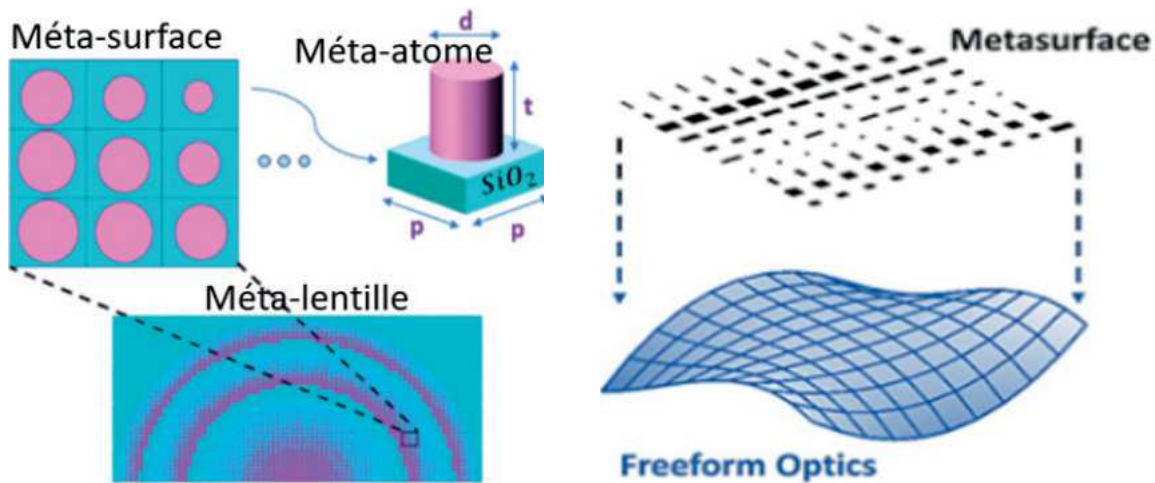


Figure 1 : a) exemple de méta-lentille tiré de [4] ; b) exemple de méta-forme tiré de [10].

QUELQUES PARAMÈTRES CLÉS

Ce sont principalement l'immersivité et la performance des dispositifs RA, en intégrant l'œil humain. Tout d'abord l'*eye-box* (*i.e.* la région 3D entre le combineur et la pupille de l'œil sur laquelle le FOV¹ est visible, à pupille donnée), le FOV (*i.e.* champ de vue sur lequel une image est projetée), mesuré en degrés et sa résolution en pixels/degrés, la MTF (fonction de transfert) du système optique formant l'image en champ proche ou lointain qui doit correspondre à la résolution de l'écran sur le FOV. Un bon système RA exige une résolution autour de 0,3 minute d'arc. Pour le commun, une résolution angulaire < 0,8 est courante (en fovéa²), mais diminue avec l'âge. Notons qu'une augmentation de la plage de distances inter-pupillaires (IPD) résulte en une augmentation de l'*eye-box*. Pour un système optique donné, ce dernier est inversement proportionnel au FOV. Il y a bien entendu d'autres paramètres. Je ne citerai ici que ceux qui jouent un rôle clé dans le dimensionnement des optiques.

LA PIERRE ANGULAIRE D'UN SYSTÈME RA : LE COMBINEUR

C'est lui qui permet la superposition du monde réel et des objets virtuels (2D ou 3D). Il est responsable de la qualité de l'imageur. Il définit la taille de l'*eye-box* et très souvent le FOV. Je ne traiterai ici que des deux principaux combineurs utilisés aujourd'hui, dans la plupart des systèmes RA.

Les combineurs optiques en espace libre

Au-delà des multiples variantes dont a fait l'objet cette configuration, il s'agit avant tout d'un dispositif d'imagerie (afficheur, optique et œil). Le combineur (Figure 2a) est la surface semi-transparente qui permet la superposition d'une image provenant d'un afficheur (*micro-display*) sur notre vision courante. Comme tout dispositif d'imagerie il

¹ <https://www.realite-virtuelle.com/fov-quest-field-view-0208/>

² La fovéa, la zone centrale de la macula, est la zone de la rétine où la vision des détails est la plus précise. Elle est située dans l'axe visuel de l'œil.

doit vérifier la condition d’aplanétisme ou des sinus d’Abbe (invariant de tout système optique stigmatique) qui relie le FOV à la résolution angulaire, et au grandissement, trois paramètres clés pour l’optimisation d’un système optique d’imagerie. Appliqué à l’imagerie RA, cela se traduit par une résolution angulaire élevée, un grand FOV et un grand *eye-box*. Or on sait que quand on élargit le FOV (*e.g.* en augmentant l’ouverture numérique du système de collimation), l’*eye-box* diminue ainsi que la résolution angulaire et par voie de conséquence, la taille de l’optique aussi.

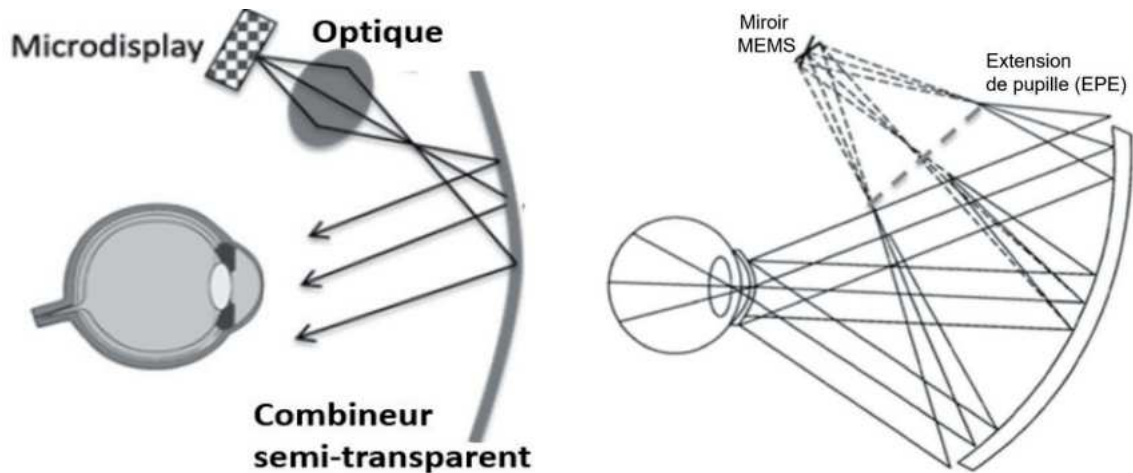


Figure 2 : a) principe du combineur tiré de [7] ; b) exemple d’extension de pupille tiré de [7].

Du fait du volume limité, on peut replier les chemins, avec le risque de créer des angles élevés et générer des aberrations supplémentaires (*e.g.* astigmatisme). Il y a plusieurs façons [7] de contourner cette difficulté : par expansion (EPE) ou duplication de la pupille (les plus courantes). La Figure 2b montre une expansion de pupille pour élargir l’IPD, condition requise pour limiter la perte de binocularité. Ceci se traduit par l’ajout d’éléments supplémentaires, avec parfois de possibles parties mobiles [12].

Ces limitations (grand angle lié souvent au repliement) démontrent que les optiques conventionnelles (réfractives) sont mal adaptées. Ceci a conduit à considérer les optiques diffractives. Nous y reviendrons plus tard. Sinon une solution moderne consiste à utiliser des *freeforms* ou méta-formes pour corriger ces forts excentremets (Figure 3a). On les retrouve dans les combineurs de type à prisme TIR (Total Internal Reflection) (Figure 3b) dont les surfaces sont souvent des *freeforms* ou méta-formes.

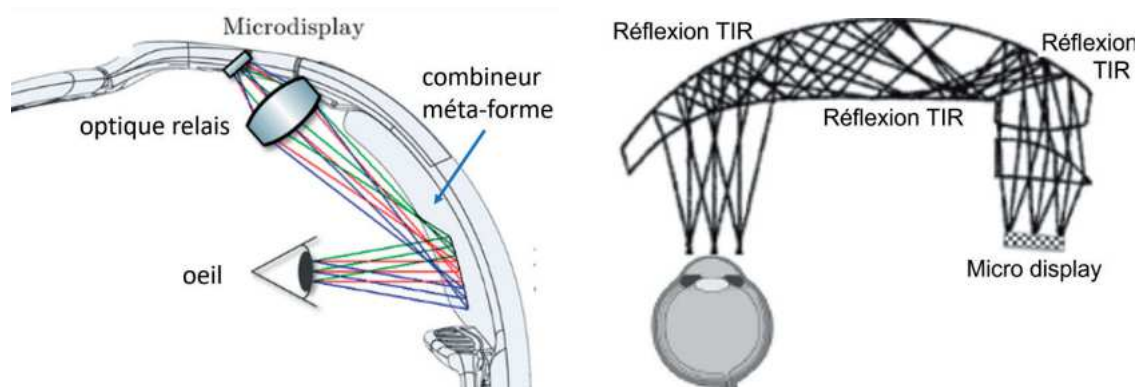


Figure 3 : a) combineur méta-forme tiré de [10] ; b) combineur prisme TIR tiré de [7].

La fabrication de *freeforms* par usinage diamant, très courante aujourd’hui, a libéré la créativité des ingénieurs pour la conception d’optiques RA, en particulier pour les guides combineurs, dans le but ouvrir l’*eye-box*, sans pénalité sur la fonction de transfert de modulation (MTF), grâce à la correction d’aberrations, combinée à l’utilisation de sources laser ou de diodes électroluminescentes, voire de diodes super-luminescentes qui permettent un champ de vision plus large (grâce à leur dispersion angulaire). Cela nous amène naturellement au second combineur.

Les combineurs optiques à base de guides de lumière

Le combineur *freeform* TIR est l’ancêtre du combineur à guide d’ondes qui repose sur la propagation de la totalité du champ dans un guide optique agissant comme un périscope [13] (Figure 4a), avec une pupille d’entrée unique et souvent plusieurs en sortie. Les soucis sont l’achromatisation du trajet et les pupilles d’entrée et de sortie, comprenant les coupleurs. Ces derniers sont l’élément clé du combineur. Ce sont des prismes, des réseaux de micro-prismes, holographiques, à relief de surface, de méta-surfaces ou de guides résonnants. On peut courber le guide, ce qui peut être avantageux. Ce combineur a fait l’objet de multiples réalisations. B. Kress y consacre un chapitre entier. J’encourage à s’y reporter. Je détaille ici une option qui constitue un champ complet de développement aujourd’hui : les coupleurs diffractifs dont on peut penser que le déploiement des méta-surfaces contribuera à imposer.

L’approche diffractive permet en effet l’emploi de surfaces minces. L’angle maximal de diffraction variant en arcsin (λ/δ), (δ est la taille de l’élément diffractant, *e.g.* le pixel du modulateur spatial – SLM), plus le pixel est petit plus le FOV sera grand. L’utilisation de méta-surfaces fixes [14] ou variables [15] permet aujourd’hui de réaliser des coupleurs de surface très compacts (*e.g.* le bloc d’entrée de la Figure 4a est supprimé). Ceci est une vraie rupture technologique. On observe une activité importante dans la conception de méta-surfaces de ce type, motivée en partie par la RA.

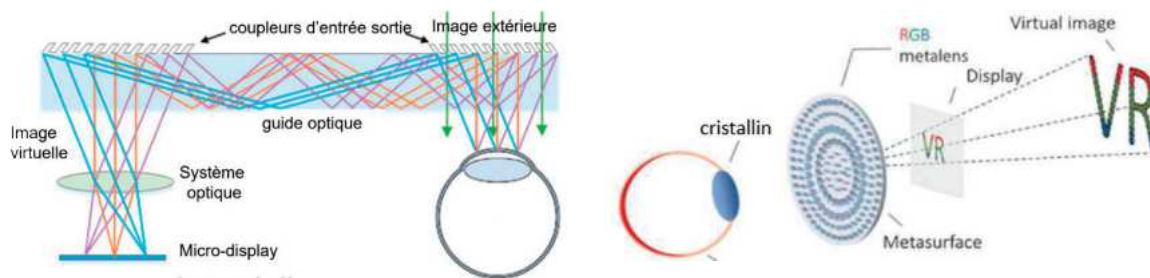


Figure 4 :

a) principe du combineur à guide tiré de [15] ; b) exemple de métalentille achromatique tiré de [16].

La Figure 4b montre une méta-surface [16] pour achromatiser la projection d’images obtenues à partir d’un écran Rouge-Vert-Bleu (RGB). Les sources sont des lasers. La conception de ce type d’éléments a beaucoup évolué. Une méta-surface n’est plus conçue de nos jours séparément des autres éléments. Connues sous le nom d’*inverse design* méta-surfaces, leur calcul intègre des connaissances *a priori* sur les images projetées (*via* une dose d’IA) pour améliorer le rendu visuel, parfois encore médiocre en imagerie holographique. Le nouvel usage de méta-atomes dits topologiques [17] est également un élément d’amélioration de l’efficacité de ces optiques. Cette pratique est indispensable au plein succès des méta-surfaces dans les combineurs RA. Ces considérations nous amènent naturellement à aborder le cas de la RA holographique.

L'APPROCHE DIFFRACTIVE HOLOGRAPHIQUE

Le terme hologramme est largement utilisé par la communauté de la RA. Il désigne les images stéréo. Il ne s'agit pas de cela ici. Ce sont généralement des éléments optiques diffractifs (DOE). Ils opèrent dans différents régimes de diffraction (Fourier, Fresnel, etc.) et sont éclairés par des sources cohérentes (*i.e.* des lasers). Des algorithmes itératifs du type IFTA [18] permettent de les calculer. L'intérêt de cette approche est la compacité du système d'imagerie et la taille du DOE. L'extension à des éléments optiques méta-surfaces (MOE) est naturelle avec une efficacité dans le visible qui reste toutefois à améliorer. Dans le domaine RA, les DOE se doivent d'être variables, ce qui requiert l'emploi de modulateurs de lumière (SLM) de phase [19] qui sont les composants clés de la RA holographique. Beaucoup de progrès ont été accomplis dans ce secteur, grâce à leur industrialisation massive pour les HUD automobiles, pour lesquels la solution diffractive est bien adaptée et compétitive [20].

Leur mise en œuvre dans les HMD est limitée, même si l'approche a l'avantage d'un faible encombrement et de grands FOV. La qualité de leur imagerie (cohérente) questionne encore pour la RA très haut de gamme. On peut parier néanmoins que le développement de la lunetterie intelligente, utilisant ce principe va accélérer leur déploiement grâce aussi à la disponibilité de lasers RGB très compacts. L'extension à des MOE dynamiques demeure complexe avec la technologie actuelle, compte tenu de la taille des méta-atomes et des variations d'indice exigées. Elle ne sera pas immédiate. Quelques solutions commencent néanmoins à voir le jour [15].

QUELQUES SOLUTIONS DE RUPTURE

La course vers plus d'immersivité conduit à concevoir des dispositifs de plus en plus proches de l'œil. Les deux exemples suivants illustrent cette démarche extrême.

Le contournement du bloc oculaire par éclairage rétinien

Une façon de s'affranchir des contraintes de dimensionnement optique consiste à exploiter la composante temporelle de la persistance rétinienne, *e.g.* par un balayage rétinien (Figure 5a). Ces dispositifs [21] existent depuis des décennies. Compacts et efficaces, ils utilisent des lasers en mode balayage ce qui permet de contourner le système oculaire en

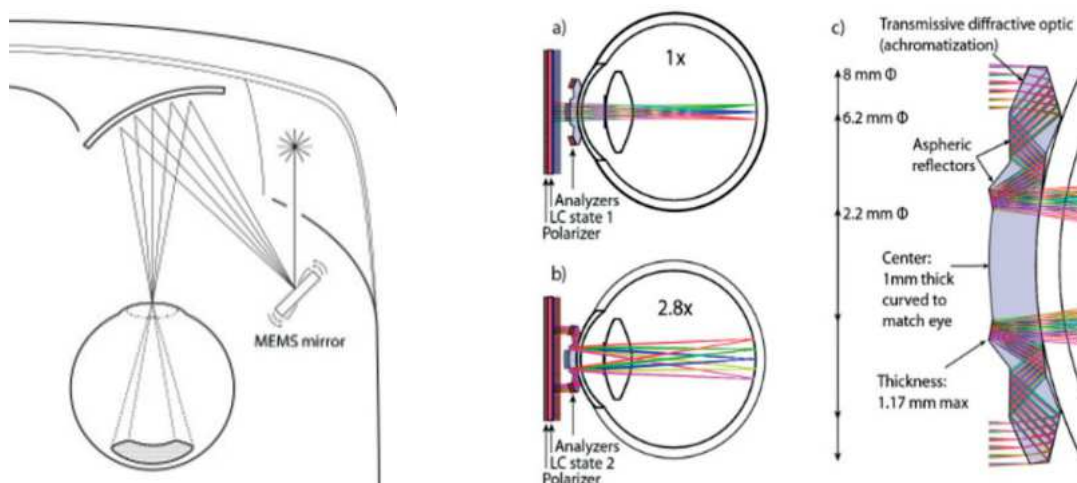


Figure 5 : a) système RA à balayage laser tiré de [7] ; b) lentille avec visée télescopique tiré de [26].

raison du faible diamètre des faisceaux pénétrant ainsi l'œil. L'image est nette et indépendante de l'accommodation ce qui évite les conflits d'accommodation-vergence. Cette solution est utilisée pour les lunettes monoculaires intelligentes ou l'affichage de textes. On est loin toutefois de la RA haut de gamme. Leur faible *eye-box* restreint le FOV effectif, et le *speckle* [22] nuit à la qualité d'image, comme la plupart des solutions diffractives. Une approche innovante contournant la limite de diffraction a été proposée récemment par le CEA-Leti [23].

Les lentilles de contact RA

Transférer une optique RA dans une lentille de contact (*i.e.* sclérale rigide) a suscité naturellement l'intérêt et les tentatives, encouragés par le succès de capteurs embarqués (*e.g.* chimiques, jauges de pression intra-oculaire [24]). La tentative RA la plus aboutie, dans les années 2020, est celle de la société californienne Mojo-Vision qui maîtrisait la technologie des afficheurs à LED [25] et qui a proposé l'intégration d'un bloc complet RA dans une lentille de contact. Malgré des avancées dans la techno d'affichage, ce projet n'a pas abouti du fait du facteur de forme trop contraint. En effet, à supposer que la surface utilisable couvre partiellement la sclère, pour une épaisseur au centre $< 500 \mu\text{m}$, le volume disponible est $< 50 \text{ mm}^3$, soit un à deux ordres de grandeur inférieurs aux facteurs de forme et de volume discutés plus haut. Outre l'obstruction de la pupille qui n'est pas souhaitable pour de multiples raisons, on voit qu'il est difficile, voire impossible, d'encapsuler dans un aussi petit volume des combineurs optiques, quelle que soit leur nature. Il existe alors deux options.

Les solutions hybrides où l'afficheur ou l'écran est hors de la lentille [26] (Figure 5b) mais assez proche pour un bon compromis FOV/*eye-box* et une encapsulation complète où l'optique de projection et la source sont dans la lentille de contact. Si dans le premier cas on retrouve le même souci d'alignement avec l'œil qu'un système RA classique, cela permet de contourner l'optique visuelle ce qui est positif. Dans l'autre cas [27], les règles d'ingénieur s'appliquent et l'approche diffractive est la plus adaptée au volume restreint [28]. Si les DOE sont fixes, une solution existe, par contre s'ils sont reconfigurables, les SLM de phase n'ont pas encore atteint la taille requise pour permettre une telle intégration, même si cela est envisageable dans le futur.

La solution PRAR [29] est un exemple de solution à DOE fixes multiples. Le dispositif comporte un ensemble de DOE disposés sur une couronne en bordure de pupille (Figure 6). Ces DOE sont en sortie d'un TIR annulaire en périphérie de la pupille et connectés à un jeu de sources laser (autant que de DOE) contrôlables à distance, éclairant ces DOE et projetant une image sur la zone rétinienne choisie.

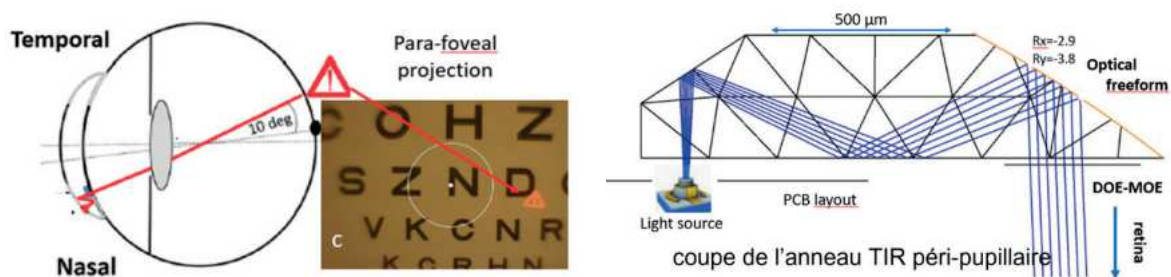


Figure 6 : a) principe de la projection parafovéale ; b) vue en coupe du TIR-PRAR tiré de [28].

Cette technique de projection pupillaire contourne le système oculaire (limitant les aberrations propres à celui-ci). Un tel TIR annulaire est facilement encapsulable. Il permet l'implantation de *freeforms* sur la surface externe réfléchissante et de DOE ou de MOE, sur la base annulaire et orientés vers l'intérieur de l'œil (Figure 6b).

QUELQUES HYPOTHÈSES ET TENDANCES

Le champ de développement des dispositifs RA est devenu très vaste. Il ne se limite plus aux aspects technologiques même s'ils en conditionnent encore largement les progrès. On voit se dessiner deux tendances marquées aux choix technologiques différents. La RA haut de gamme avec écrans/afficheurs à grande étendue, pour des contenus 3D de haute qualité. Ceux-ci ont bénéficié des avancées du surfaçage *freeform* qui s'intensifieront avec les méta-formes. Ces améliorations incrémentales devraient surtout impacter les guides/combineurs. L'emploi d'optiques méta minces à grande ouverture devrait permettre un gain significatif en immersivité.

La voie holographique, qui va profiter pleinement des lasers bas coût miniaturisés, va gagner en qualité et bénéficier du déploiement de la lunetterie intelligente moins exigeante en RA-3D. On sera davantage dans la version *consumer* d'un HUD, pour lequel cette approche a de nombreux atouts. L'impact du surfaçage *freeform* est moindre ici. L'enjeu est plutôt côté méta-lentilles fixes et surtout reconfigurables, qui constituent encore un véritable défi technologique.

En matière de lentilles de contact, un dispositif RA-3D riche, dans une lentille de contact est très difficile à imaginer, du fait du volume très restreint. La solution diffractive/holographique est par contre envisageable, à partir d'éléments fixes projetant des images directement sur la rétine. On peut imaginer ces solutions pour les systèmes de vigilance ou de navigation, et surtout pour des dispositifs médicaux (*e.g.* assistance et réhabilitation visuelle). L'extension à une version dynamique impose une miniaturisation importante des SLM, ce qui n'est pas pour tout de suite.

Une ligne de partage semble se dessiner entre RA riche haut de gamme et RA holographique pour des dispositifs plus intégrés comme des lunettes, voire des lentilles de contact. Cette distinction est sans doute temporaire, car conditionnée aux progrès technologiques (*e.g.* miniaturisation des écrans) et au déploiement à l'échelle industrielle de méta-surfaces, comme l'a montré STMicroelectronics avec les caméras de *smartphone*.

BIBLIOGRAPHIE

- [1] COLBURN W. S. *et al.* (1977), "Holographic combiners for head up displays", Tech Report, n°AFAL-TR-77-110.
- [2] BARFIELD W. (2015), *Fundamentals of wearable computers and augmented reality*, Second Edition, Taylor & Francis, Boca Raton, FL.
- [3] AMITAI Y. *et al.* (1989), "Holographic elements with high efficiency and low aberrations for helmet displays", *Applied Optics*, 28, pp. 3405-3416.
- [4] LEE C. *et al.* (2022), "Tutorial on metalenses for advanced flat optics: design, fabrication & critical considerations", *Journal of Applied Physics*, DOI :10.1063/5.0078804.
- [5] <https://newsroom.st.com/media-center/press-item.html/t4458.html>
- [6] <https://www.usine-digitale.fr/article/meta-et-essilorluxottica-lancent-leurs-nouvelles-lunettes-connectees-ray-ban-meta.N2175757>
- [7] KRESS B. (2020), *Optical architectures for augmented-, virtual-, and mixed-reality headsets*, SPIE Press, Bellingham, Washington, January.
- [8] <https://centerfreeformoptics.org/>
- [9] THOMPSON K. P. *et al.* (2012), "Freeform optical surfaces: a revolution in imaging optical design", *Optics and Photonics News*, vol. 23, Issue 6, pp. 30-35.

- [10] NIKOLOV D.K. *et al.* (2021), “Metaform optics: bridging nanophotonics & freeform optics”, *Science Advance*, vol 7, Issue 18, DOI : 10.1126/SCIADV.ABE5112.
- [11] BENISTY H., GREFFET J-Y & LALANNE PH. (2024), *Introduction to nanophotonics*, Oxford University Press.
- [12] McMANAMON P. F. *et al.* (2009), “A review of phased array steering for narrow-band electrooptical systems”, *Proc. IEEE*, 97(6), pp. 1078-1096.
- [13] GOPAKUMAR M. *et al.* (2024), “Full-color 3D holographic augmented reality displays with metasurface wave guides”, *Nature*, 629(8013), pp. 1-7.
- [14] ZHUO W. *et al.* (2020), “A review of high-efficiency Pancharatnam-Berry metasurfaces”, *Terahertz Science & Technology*, vol. 13, n°3, pp. 73-89.
- [15] DOLAN J. *et al.* (2021), “Broadband liquid crystal tunable metasurfaces in the visible: liquid crystal inhomogeneities across the metasurface parameter space”, *ACS Photonics*, 8(2), DOI : 10.1021/acsp Photonics.0c01599.
- [16] PARK Y.-W. *et al.* (2021), “Meta-optics achieves RGB achromatic focusing for virtual reality”, *Sciences Adv.*, 7, eabe4458, DOI :10.1126/sciadv.abe4458.
- [17] LI Z. *et al.* (2022), “Empowering metasurfaces with inverse design: principle and applications”, *ACS Photonics*, vol. 9, Issue 7, pp. 2178-2192.
- [18] KRESS B. & MEYRUEIS P. (2007), *Applied digital optics: from microoptics to nanophotonics*, 1st edition, John Wiley and Sons, New-York.
- [19] Par exemple voir : <https://holoeye.com/>
- [20] HE X. (2024), “Holographic projections: the future of automotive heads-up displays”, IDTechEx Reports, April 17.
- [21] LIN J. *et al.* (2016), “Retinal projection head-mounted display”, *Frontier of Optoelectron*, DOI : 10.1007/s12200-016-0662-8.
- [22] YANG X. *et al.* (2019), “A fast computer-generated holographic method for VR and AR near-eye 3D display”, *Applied Science*, 9i (19), 4164, doi.org/10.3390/app9194164.
- [23] MARTINEZ C. *et al.* (2023), “Sparse holographic imaging for an integrated AR near-eye display”, *Applied Optics*, vol. 62, n°8, March.
- [24] ZHANG J. *et al.* (2022), “Smart soft contact lenses for continuous 24-hour monitoring of intraocular pressure in glaucoma care”, *Nature Com*, vol. 13, n°5518.
- [25] ZHANG S. *et al.* (2023), “Research progress of micro-LED display technology”, *Crystals*, 13(7), 1001, <https://doi.org/10.3390/cryst13071001>.
- [26] TREMBLAY E. J. *et al.* (2013), “Switchable telescopic contact lens”, *Optics Express*, vol. 21, Issue 13, pp. 15980-15986.
- [27] KHALDI A. *et al.* (2020), “The cyclops contact lens: a laser emitting contact lens for eye tracking”, *Nature Scientific Report*, vol. 10, 14804.
- [28] DE BOUGRENET DE LA TOCNAYE J.-L. (2024), “Novel smart contact lens eye assistants”, Invited speaker at Gordon Research Conference Image Science, Maine, June.
- [29] NOURRIT V. *et al.* (2021), “Perifoveal retinal augmented reality (PRAR) on contact lenses”, *Optical Engineering*, 60(11), 115102, <https://doi.org/10.1117/1.OE.60.11.115102>

Métavers : au-delà des « casques de réalité virtuelle »

Par Guillaume MOREAU

Directeur délégué à la Recherche et à l'Innovation à l'IMT Atlantique

Depuis le discours de Mark Zuckerberg de 2021 sur le « métavers », force est de constater que la généralisation annoncée n'a pas encore eu lieu. Si les technologies immersives ont indéniablement progressé et atteint des niveaux de prix plus compatibles avec le grand public, certains défis, plus ou moins liés aux compromis effectués sur les interfaces, restent à relever. Non seulement les logiciels doivent s'adapter au compromis masse, énergie, bande passante réseau, capacité de calcul, mais les outils auteurs doivent également progresser. L'ensemble de ces systèmes n'atteint pas les capacités de la perception humaine, engendrant des effets secondaires et potentiellement des conséquences sanitaires. Enfin, il convient d'étudier les enjeux éthiques de ces technologies qui se marient progressivement à l'intelligence artificielle.

Depuis l'annonce de Mark Zuckerberg sur le « métavers » en novembre 2021, la tendance s'est retournée et la généralisation de la réalité virtuelle auprès du grand public n'a pas encore eu lieu. C'est donc l'occasion de faire un point d'étape sur les grands défis qui restent à résoudre en vue d'une adoption plus massive.

Tout d'abord, qu'entendait-il par « métavers » ? Il s'agissait de la combinaison d'une réalité virtuelle (un monde virtuel appelé Horizon Worlds), des moyens d'accès à ce monde virtuel, de possibilités d'interactions avec d'autres humains et quelques éléments autour de la monétisation. Par réalité virtuelle, on entend ici le fait de s'immerger, par exemple avec un casque comme le Meta Quest, dans un environnement artificiel pouvant représenter en partie la réalité et d'interagir en temps réel avec ce monde virtuel (Arnaldi *et al.*, 2018). On peut par exemple interagir avec des représentations d'êtres humains qu'on appelle alors des avatars, correspondant à des personnes réelles ou potentiellement mus par l'intelligence artificielle. La réalité mixte est une forme de généralisation de la réalité virtuelle où l'utilisateur perçoit simultanément des éléments du monde réel et des éléments de synthèse. Les technologies permettant d'accéder à cette réalité virtuelle font l'objet d'un autre article dans ce numéro, nous nous concentrerons sur les autres aspects. Parmi ceux-ci, il y a bien sûr des considérations logicielles liées aux nécessaires compromis matériels, à la représentation des mondes virtuels et en particulier des avatars mais aussi aux capacités du système sensori-moteur et cognitif humain. Enfin, qui dit enjeux humains dit nécessairement enjeux éthiques.

C'est ce que nous nous proposons d'aborder dans la suite de cet article.

ENJEUX LOGICIELS

Même si les matériels progressent, ils relèvent nécessairement d'un compromis entre leurs performances techniques, leur encombrement et leur poids. Tous ces paramètres

influent sur leurs capacités de calcul, de communication et leur autonomie énergétique. Se pose également la question de leurs moyens d'interaction avec l'environnement et de la création des mondes artificiels, y compris pour le cas particulier des avatars.

Du point de vue informatique, le rendu d'une scène en réalité virtuelle ressemble de très près à la synthèse d'images en temps réel telle qu'elle est pratiquée dans les jeux vidéo : à partir d'un modèle 3D de la scène généralement représenté sous forme de formes géométriques simples comme des triangles, d'un modèle d'éclairage, d'un modèle de caméra virtuelle et de lois simplifiées du rendu (modèle de Phong), on calcule le rendu graphique « suffisamment vite » (25 images par seconde sont un minimum pour la rémanence rétinienne, mais 60-90 images par seconde semblent constituer des valeurs limitant la cybercinétose). La différence se situe dans les paramètres de la caméra et le calcul de sa « pose », *i.e.* sa position et son orientation. Les casques modernes sont dotés de capteurs permettant ce calcul moyennant un algorithme de fusion de données. Le problème est bien plus complexe en réalité augmentée, comme on le verra plus loin. Ces calculs peuvent être effectués à différents niveaux : directement sur le casque ce qui limite le besoin en bande passante réseau au détriment de la consommation énergétique, sur un PC avec carte graphique dédiée, ou carrément dans le « nuage ». On retrouve les problématiques habituelles des systèmes embarqués entre technologies avec et sans fil, sécurité, bande passante et autonomie énergétique.

La création de monde virtuels est également plus complexe que la création de monde 3D pour l'image de synthèse ou l'animation (cinéma, jeux vidéo) : elle nécessite la création de moyens d'interaction entre le ou les utilisateurs et les différents éléments du monde afin de sélectionner, manipuler ou modifier les objets du monde virtuel. C'est un domaine de recherche très actif entre extension des paradigmes 2D comme l'utilisation de la souris en 3D à des expériences très diversifiées comme le suivi de la direction du regard pour la sélection en 3D (Paulus, 2021), la détection de gestes, la commande vocale... À l'inverse, comme il est quasi-impossible d'utiliser simultanément un clavier et un casque de RV, l'entrée de texte pour annoter un monde virtuel reste extrêmement peu efficace quelle que soit la métaphore proposée.

La combinaison du monde virtuel et des différents scénarios d'interaction qui était directement programmée par-dessus des *middlewares* se limitant à une couche d'interface matériel-logiciel est aujourd'hui à peu près intégrée dans un outil devenu un standard de fait, comme Unity 3D. Il reste néanmoins à créer et animer des avatars : il faut un modèle géométrique (basique ou visuellement réaliste, ce que les futurs outils de Meta se promettent de générer à partir d'un simple *smartphone*), un modèle d'animation en lien avec la biomécanique humaine et surtout un modèle comportemental. Ce dernier qui a fait l'objet de recherches actives depuis les années 1990, avec un beau succès dans le domaine de la simulation de foules, se construit aujourd'hui avec les techniques d'intelligence artificielle. Tous ces types de modèles se trouvent aujourd'hui en ligne, mais ils doivent être adaptés : par exemple, on ne peut pas simplement transformer un modèle d'humain virtuel de 1,65 m en un modèle d'1,90 m, qui ne marchera pas de la même manière. Il faut respecter certaines contraintes biomécaniques, c'est l'objectif des techniques de "motion retargeting". Certains peuvent aussi souhaiter un avatar qui ne leur ressemble pas.

Le dernier élément régulièrement évoqué est celui du couplage entre avatars réalistes et intelligence artificielle. Du point de vue technique, cela implique d'importants coûts de calcul qui ne peuvent pas être réalisés aujourd'hui sur un dispositif portable et ont pour conséquence des besoins en bande passante. L'IA embarquée est un important défi impliquant des travaux sur les architectures des composants, la compression de réseaux de neurones... La crédibilité d'un avatar intégralement piloté par intelligence artificielle n'est pas acquise aujourd'hui et soulève des enjeux éthiques que nous étudierons plus loin.



Figure 1 : Exemple de simulation de foule virtuelle (Source : Inria).

CAS PARTICULIER : LA RÉALITÉ AUGMENTÉE

En matière de réalité augmentée, la problématique du calcul est bien plus importante parce qu'il faut faire « coïncider » le monde réel et le monde virtuel : si un objet virtuel est posé sur une table réelle, il ne doit pas apparaître ni au-dessus ni en dessous de celle-ci, ni encore donner l'impression de trembler autour de celle-ci. Le calcul précis et fréquent de la position et de l'orientation du casque est donc une condition nécessaire.

En intérieur, les systèmes incorporés comme celui l'HoloLens 2 font appel à des techniques complexes mêlant projection de lumière structurée, caméras infrarouges et optiques, SLAM et fusion de données qui fonctionnent dans un espace de travail de 5 à 10 m de côté. En extérieur, le problème se corse puisque les caméras de profondeur utilisant l'infrarouge sont inefficaces. Les systèmes GNSS ne fournissent qu'une position de précision décimétrique, pas d'orientation et fonctionnent à une fréquence d'environ 1 Hz. Cela ne résout pas en outre le fait que les systèmes de réalité augmentée actuels ne sont pas suffisamment lumineux pour permettre d'afficher une information dans un environnement ensoleillé.

On retrouve cette question d'éclairage dans la scène mixte : il faut respecter les interactions entre les objets réels et les objets virtuels. En premier lieu, il convient de respecter les occultations entre les objets ; techniquement, l'écran d'un équipement de réalité augmentée est placé entre l'œil et le monde réel, les objets virtuels seront donc toujours devant les objets réels. Comme on ne peut pas altérer le monde réel, il faut alors calculer les interactions entre les différents objets réels et virtuels *via* une technique de "phantom rendering" qui suppose l'acquisition d'un modèle 3D de la scène réelle (Kalkofen *et al.*, 2011). Les techniques de modélisation 3D d'une scène réelle existent mais sont difficilement exploitables en temps réel lorsque la scène est dynamique. Une reconstruction purement géométrique de la scène n'est pas suffisante, il faut y ajouter une reconstruction photométrique afin d'assurer la cohérence de l'éclairage des deux parties de la scène. En sortie de cette chaîne, nous arrivons donc à une image virtuelle calculée qu'il faut afficher, ce qui renvoie à un problème de matériel : l'immense majorité des systèmes de réalité augmentée actuels ne savent qu'ajouter de la lumière à un pixel donné superposé au monde réel. Pas pratique pour montrer l'ombre d'un objet virtuel projetée sur un objet réel !

ENJEUX HUMAINS

Perception

La conséquence première de l'imperfection de ces dispositifs matériels et logiciels est l'impact sur le système perceptif humain. Si on compare les performances d'un casque de réalité virtuelle (par exemple un Meta Quest 3) au système visuel humain, le champ de vision horizontal atteint aujourd'hui 100 à 120° (au lieu de 180°), l'acuité visuelle monoscopique est à peu près du même ordre mais il subsiste un problème fondamental : en vision naturelle, lorsque l'humain regarde un objet, l'œil accommode sur l'objet et les deux yeux convergent également sur l'objet d'intérêt. Dans un dispositif stéréoscopique de type casque ou CAVE, les yeux accommodent toujours sur l'écran mais convergent sur l'objet d'intérêt qui peut être dans un plan différent de celui de l'écran. Cette rupture de la relation entre accommodation et convergence n'est pas naturelle (et d'ailleurs pas toujours possible). De plus, en vision naturelle, les objets qui ne sont pas autour de la zone d'intérêt ne sont pas fusionnés (diplopie) et sont flous. En vision stéréoscopique, faute de savoir suffisamment rapidement ce que les yeux regardent, il n'est pas possible de flouter en temps réel les zones non fusionnées, ce qui engendre une fatigue visuelle.

L'ensemble de ces imperfections engendrent des différences de perception entre le monde réel et le monde virtuel. Typiquement, celui-ci est perçu comme étant de taille différente (Peillard, 2020).

Qui plus est, la qualité de reproduction des *stimuli* sur l'ensemble du système sensoriel est différente selon les modalités sensorielles, le contenu de l'application et les dispositifs techniques retenus, engendrant des incohérences sensorielles pouvant être source d'inconfort mais aussi de limitation de la plausibilité ou du sentiment de présence dans le monde virtuel. Cependant, le lien entre complétude de la perception et présence est loin d'être établi : si un minimum de qualité des *stimuli* est nécessaire pour engendrer un sentiment de présence, la théorie de la « vallée de l'étrange » de Mori (2012) tend à l'inverse à dire que la perception se focalisera sur les différences les plus ténues.

Effets sanitaires de la réalité virtuelle

Tout dispositif de réalité virtuelle, artificiel par nature, sollicite le système sensori-moteur et donc cognitif de l'être humain qui l'utilise. Notamment en raison de performances encore éloignées du système perceptif humain, il n'est pas exempt d'effets sur la santé humaine, ce que démontre un rapport de l'ANSES de 2021 en séparant ces effets en fonction des niveaux de preuve (Burkhardt *et al.*, 2021).

Les effets avérés, dont l'intensité dépend des interfaces utilisées, du contenu proposé mais aussi de la sensibilité individuelle, sont :

- la cybercinétose (30 à 50 % des utilisateurs) qui peut avoir comme symptômes des nausées, des vomissements, des maux de tête ainsi que divers autres effets physiologiques ou vestibulaires : la cause en reste mal connue, les incongruences sensorielles restent les plus citées ;
- l'altération de la coordination sensori-motrice et des capacités perceptives après exposition ;
- les effets liés au fonctionnement des casques en tant que dispositifs émetteurs de rayonnement : lumière bleue qui perturbe le cycle circadien et le sommeil en particulier chez les enfants, modulation temporelle de la lumière qui a un effet sur le déclenchement de crises chez les personnes souffrant d'épilepsie et engendre de la fatigue visuelle.

Le rapport souligne aussi des effets possibles sur lesquels il n'est pas aujourd'hui possible de conclure en raison d'un niveau de preuve encore insuffisant :

- les effets liés aux interfaces elles-mêmes et à leur ergonomie, incluant les accidents (collisions avec l'environnement par distraction ou occultation), des troubles musculosquelettiques dus par exemple au poids des casques et à leur répartition des masses ;
- les effets psychologiques et psycho-sociaux à l'instar des jeux vidéo (risques émotionnels, déréalisation, dépendance, isolement social...). Ces risques sont naturellement plus liés aux contenus qu'aux dispositifs de réalité virtuelle ; ces risques induisent de façon indirecte des risques sur le développement émotionnel et cognitif ;
- les effets sur la représentation de soi dans le monde virtuel sur lesquels nous reviendrons dans le paragraphe consacré à l'éthique.

Enjeux éthiques

Un récent rapport du CNPEN (2024) pointe les risques liés au « métavers » tout en mentionnant la difficulté d'étudier un ensemble de technologies à vocation intégrative, elles-mêmes en cours de développement.

Si certains pointent à raison que le métavers peut être un moyen d'économiser un certain nombre de déplacements, l'équilibre écologique du métavers et des technologies associées est loin d'être une évidence : les moyens opto-électroniques sont conséquents, le cycle de vie avant obsolescence est aujourd'hui très court et si la consommation électrique des casques reste par définition limitée, il n'est pas de même pour l'infrastructure de calcul (rendu, IA), de stockage et de communication.

Le métavers est-il inclusif ? Il y a plusieurs moyens de répondre à cette question : d'un côté, malgré la baisse sensible des coûts d'accès aux dispositifs et à Internet, il est bien évident que tout le monde n'est pas égal devant l'accès à ces technologies, notamment mais pas seulement dans les pays du Sud. À l'opposé, ces inégalités de richesse essentiellement peuvent être partiellement compensées grâce au métavers : un concert ou un événement sportif dont le nombre de places est limité devient accessible à plus de monde, un site historique situé à des heures d'avion ou fermé à cause du surtourisme est à nouveau accessible, même s'il faut rester conscient que l'expérience est dégradée.

Certains points sont d'un abord plus complexe, notamment en ce qui concerne les avatars : ces représentations des êtres humains, de plus en plus précises et « convaincantes » soulèvent quelques questions importantes :

- L'identification des personnes dans le métavers : si ressembler à autre chose que son corps physique peut être souhaitable, on ne peut pas négliger la possibilité d'usurpation d'identité très convaincante (*deepfake*) avec ses conséquences possibles : influence, vol de données, abus de confiance...
- De façon plus subtile encore, la littérature scientifique commence à démontrer que la modification de la représentation de son propre corps est à l'origine de modifications comportementales : être incarné dans un corps de couleur tend par exemple à diminuer les préjugés racistes (Peck *et al.*, 2013), tandis qu'une incarnation dans un corps de super-héros rend l'utilisateur plus susceptible d'aider son prochain. On peut se demander ce qu'une incarnation dans un corps représentant quelque chose de moins positif pourrait donner.
- Dans tous les cas, la notion de responsabilité doit être établie. Si le vol de données est qualifié aujourd'hui, le vol de biens virtuels n'est pas complètement défini et certaines notions restent à compléter : il y a d'ores et déjà quelques ressentis de violences sexistes et sexuelles voire de viols dans le métavers qui vont au déjà de la définition

légale du viol du code pénal. Pour autant, ceci est en contradiction avec l'objectif d'un certain nombre de jeux vidéo, moyen clé d'accès au métavers où tuer d'autres êtres « vivants » est un objectif en soi, qu'ils soient animés par le moteur de jeu ou des représentations d'autres êtres vivants. Deux systèmes moraux sont amenés à cohabiter.

CONCLUSION

Les technologies immersives sont d'ores et déjà utilisées dans de grands nombres de domaines de l'industrie, du monde de la santé et du divertissement. La généralisation est lente faute d'application incontournable (*silver bullet*) même si des défis techniques restent à régler. Pour autant, on voit progressivement les outils de réalité virtuelle passer du praticien au patient dans le domaine médical. La réalité virtuelle a des bénéfices réels et prouvés pour le traitement de certaines pathologies comportementales comme les phobies (Silva Freitas *et al.*, 2021). Elle propose néanmoins aujourd'hui une expérience dégradée par rapport à la réalité et si nous faisons une analogie avec le fameux test de Turing, le mélange ultime IA-RV ne passerait pas ce test et ne le passera probablement jamais complètement pour une personne avertie. La généralisation nécessite que le législateur se penche sur les vides juridiques qui subsistent au-delà de l'AI Act, mais gageons que nous saurons raisonner en termes de bénéfices – risques à l'instar de ce qui se fait en thérapie pour ne pas tuer dans l'œuf une technologie qui représente aussi une immense opportunité pour proposer des outils inclusifs au service de l'ensemble de la société et en particulier auprès des populations privées de capacités physiques ou cognitives, financières.

BIBLIOGRAPHIE

ARNALDI B., GUITTON P. & MOREAU G. (dir.) (2018), *Réalité virtuelle et réalité augmentée : mythes et réalités*, Londres, ISTE.

BURKHARDT J.M., BEHAR-COHEN F. GRYSZPAN O. KLINGER E., LOBJOIS R., MOREAU G. NANNIPIERI O., PALJIC A. PIOLINO P., HUNG T.V., TISSERON S. VIAUD-DELMON I., DAVIDOVICI M., ATTIA D. & BAYEUX T. (2021), « Expositions aux technologies de réalité virtuelle et/ou augmentée », Paris, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.

CNPEN (2024), « Avis n°9 Métavers : enjeux d'éthique », Paris, Comité national pilote d'éthique du numérique.

KALKOFEN D., SANDOR. C, WHITE S. & SCHMALSTIEG D. (2011), "Visualization techniques for augmented reality", *In: Handbook of Augmented Reality*, pp. 65-68.

MORI M., MACDORMAN K. & KAGEKI N. (2012), "The uncanny valley" (traduction), *IEEE Robotics & Automation Magazine*, 19(2), pp. 98-100.

PAULUS Y. T. & REMIJN G. B. (2021), "Usability of various dwell times for eye-gaze-based object selection with eye tracking", *Displays*, 67.

PECK T.C., SEINFELD S., AGLIOTI S.M. & SLATER M. (2013), "Putting yourself in the skin of a black avatar reduces implicit racial bias", *Consciousness and Cognition*, 22(3), pp. 779-787.

PEILLARD E. (2020), *Vers une caractérisation des biais perceptifs en réalité mixte : une étude de facteurs altérant la perception des distances*, thèse de doctorat, École Centrale de Nantes.

SILVA FREITAS J.R., SILVA VELOSA V.H., NUNES ABREU L.T., JARDIM R.L., VIEIRA SANTOS J.A., PERES B. & CAMPOS P.F. (2021), "Virtual reality exposure treatment in phobias: a systematic review", *Psychiatric Quarterly*, 92, pp. 1685-1710.

La cybersécurité

Par Hervé DEBAR

Professeur de l'Institut Mines-Télécom à Télécom SudParis

Les enjeux de cybersécurité sont devenus une problématique majeure des infrastructures numériques. La menace portée par des attaquants organisés, outillés et motivés ne laisse aucun doute sur le fait que ces infrastructures numériques sont sous attaque et que certaines seront compromises. Les méthodes et outils de la cybersécurité doivent donc être pris en compte dans la spécification et le déploiement de ces infrastructures numériques actuelles et futures, pour analyser le risque et l'impact des attaques, pour mettre en place des mécanismes de protection, de détection et de remédiation face à ces attaques. Nous avons fait des progrès ces dernières années pour contenir les attaques informatiques, mais il reste absolument nécessaire d'inclure les problématiques de cybersécurité en continu dans le pilotage de ces infrastructures numériques. Cette prise en compte des problématiques de sécurité participera naturellement d'un fonctionnement plus efficace des services offerts aux utilisateurs de ces infrastructures.

INTRODUCTION

Les infrastructures numériques sont présentes dans tous les services que nous utilisons de manière quotidienne. De ce fait, ces infrastructures numériques sont devenues partie intégrante de services critiques et on trouve du code informatique et des besoins en communication dans de nombreux environnements. Ces outils sont utilisés d'une part pour capter et traiter des informations, d'autre part pour des infrastructures de commande et de contrôle. La collecte et l'acheminement des informations nécessaires au bon fonctionnement de ces infrastructures reposent donc en grande partie sur une connectivité ubiquitaire, donc sur des réseaux offrant des capacités de communication et de traitement d'une grande diversité.

POURQUOI LA CYBERSÉCURITÉ DES RÉSEAUX EST-ELLE UN ENJEU MAJEUR ?

Tout d'abord, il est nécessaire de mentionner que le monde numérique est de plus en plus connecté et interconnecté, et ne peut fonctionner sans ces interconnexions. Par conséquent, l'approche par isolation (suppression de la connectivité) totale ou partielle est mise en difficulté par l'absence de périmètre à isoler (réseaux sans fil par exemple) ou par la difficulté de mettre en place des outils de filtrage des communications efficaces (déploiement très large du chiffrement de bout en bout, complexité d'analyser le contenu des données échangées plutôt que des métadonnées ou des enveloppes).

Dans la mesure où il devient difficile d'identifier et de localiser un objet numérique ou son propriétaire et son usage légitime, il devient de plus en plus difficile de mettre en place des politiques de sécurité pouvant répondre aux besoins croissant de protection de ces infrastructures numériques.

Un enjeu économique

Le coût global de la cybercriminalité est estimé à 7,08 T\$ en 2022 et pourrait doubler en 6 ans¹.

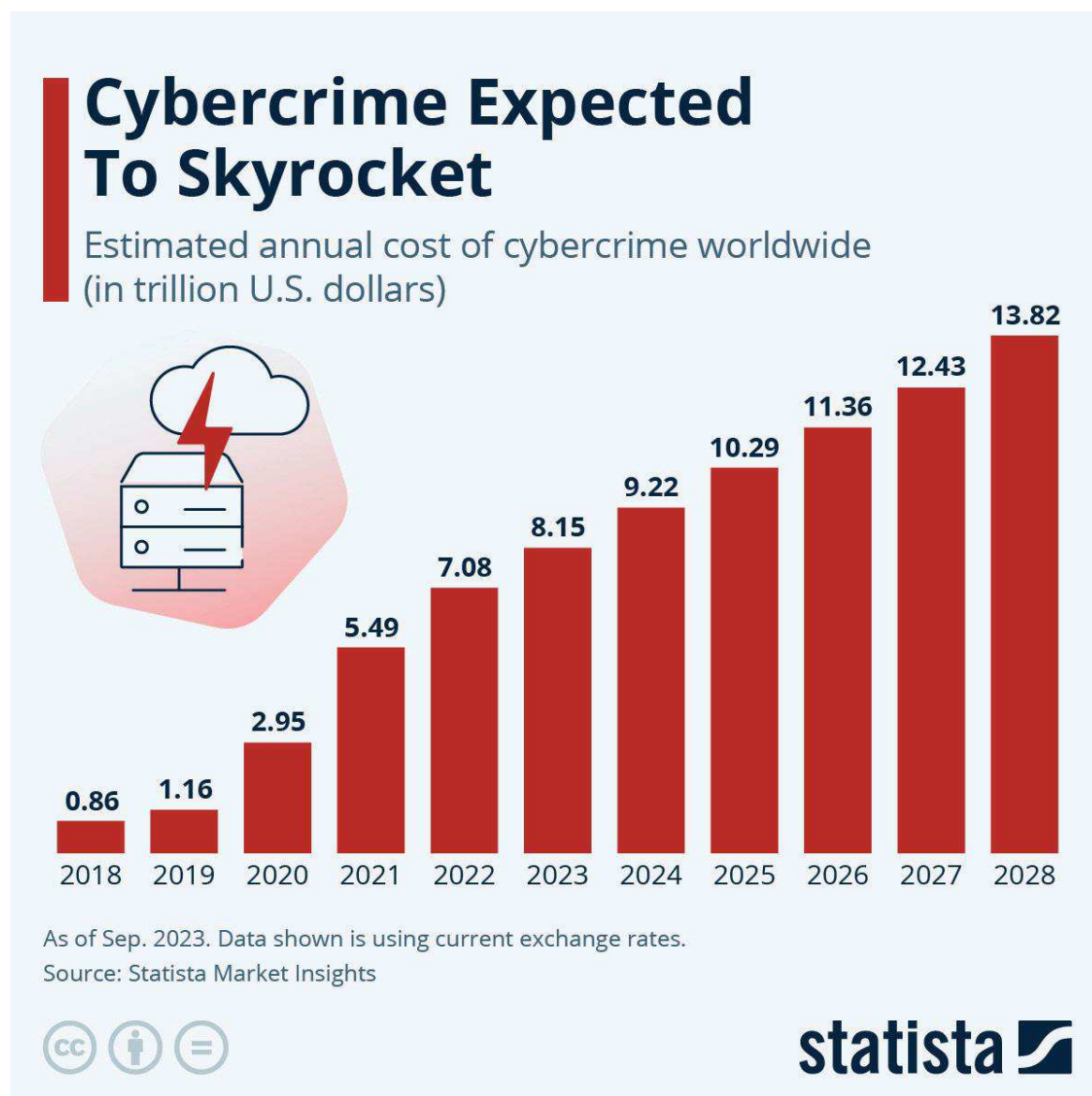


Figure 1 : La montée de la cybercriminalité, en trillions US dollars (Source : Statista Technology Market Outlook ; National Cyber Security Organizations, FBI, IMF).

Il est nécessaire de comprendre que les attaques cyber ne constituent plus un épiphénomène, mais une source de perte de valeur importante pour l'ensemble de l'économie. Le cybercrime est devenu une source de revenu majeure pour les cybercriminels (plus que le trafic de drogue), et il est particulièrement attrayant grâce aux attaques à distance. Il est en effet difficile de traquer les criminels et d'amasser les preuves nécessaires à une condamnation, et cette activité illégale est sensiblement moins risquée.

Même si les chiffres sont difficiles à établir, les dégâts d'une attaque cyber réussie sont très importants et prennent des formes multiples, amenant à une indisponibilité de fonctionne-

¹ <https://www.weforum.org/agenda/2024/01/cybersecurity-cybercrime-system-safety/>

ment de l'organisation, à une perte de chiffre d'affaires et *in fine* à une perte de confiance des clients. Les coûts des mécanismes de protection sont également significatifs, et ils ne garantissent pas une protection absolue. Cependant, pour un grand nombre d'attaques, ils permettent d'empêcher la compromission ou d'en limiter les effets et la durée.

Notons également que les compromissions ne sont pas immédiatement détectées, même si la situation dans ce domaine s'améliore².

La Figure 2 montre qu'en 10 ans, nous avons divisé par 20 le temps qu'un attaquant passe en moyenne dans un système compromis avant d'être détecté. Le même rapport indique toutefois que la détection se fait le plus souvent par des sources externes alertant l'organisation d'une compromission. Ce temps pendant lequel un attaquant reste résidant dans un système d'information lui permet largement de cartographier le système et d'en extraire des informations sensibles, ce qui induit des dommages significatifs.



Figure 2 : Durée moyenne avant qu'une compromission soit détectée dans un système (Source : M-Trends 2023).

Attaques à grande échelle *versus* attaquants motivés

La connectivité a été un facteur majeur dans le développement des attaques cyber. On recense ainsi le premier vers informatique en 1988, ainsi que de premiers actes d'espionnage (The Cuckoo's Egg, 1988). Les réseaux, et la capacité de se connecter à distance, ont joué un rôle majeur, et ont créé des communautés de *hackers* souhaitant accéder à des infrastructures auxquelles ils n'avaient pas accès facilement.

Le profil des attaquants a cependant sensiblement évolué depuis les débuts d'Internet. Nous trouvons aujourd'hui deux grandes familles d'attaquants, catégorisés par leur motivation :

- Motivation économique : les attaquants guidés par un motif économique souhaitent tout simplement gagner de l'argent. Ils utilisent des outils disponibles à faible coût, louent leurs infrastructures, et se basent sur des attaques à grande échelle (par exemple des campagnes de *phishing*) pour toucher le plus de victimes potentielles possible. Même si leur taux de succès est faible, le faible coût d'investissement et le passage à l'échelle suffisent à rendre ces opérations rentables. L'usage de l'intelli-

² <https://www.mandiant.fr/resources/blog/m-trends-2023>

gence artificielle permet également de rendre ces opérations de plus en plus crédibles et difficiles à différencier d'opérations légitimes. Ces attaquants sont le plus souvent des criminels, individus ou groupes.

Cette motivation économique peut amener également à des destructions de données ou des destructions physiques d'équipements, causant des dommages majeurs aux victimes. Ces dommages peuvent être intentionnels, et toucher le monde physique, comme le montrent les attaques contre les hôpitaux et les objets connectés médicaux. Il est également possible de prendre le contrôle à distance d'objets (par exemple un véhicule) pouvant devenir une arme par destination.

- Motivation liée à la cible : certains attaquants choisissent une cible particulière et déploient de grands efforts pour compromettre cette cible. Ils disposent de beaucoup de capacités (financières, techniques, humaines) pour mener à bien leurs attaques. Leurs buts peuvent être parfois économiques, mais sont le plus souvent liés à de l'espionnage ou de la désinformation. Ces attaquants sont le plus souvent soutenus par des États-nations, soit directement (armée), soit indirectement (terrorisme, mafias).

Dans les deux cas, force est de constater qu'il est très difficile d'être complètement protégé des attaques, et qu'il est devenu indispensable de pratiquer une défense en profondeur pour limiter les dégâts.

Infrastructures et données

La cybersécurité s'applique à deux objets, les infrastructures numériques et les données qu'elles contiennent. Il est donc indispensable de penser sa cybersécurité suivant ces deux dimensions.

La cybersécurité des infrastructures numériques consiste à déployer des mécanismes de gestion du fonctionnement de ces infrastructures permettant de gérer le risque. Il s'agit donc, de manière très générale, de s'assurer que le fonctionnement de l'infrastructure est conforme aux règles décidées par l'organisation, et par ses régulateurs. Cela repose sur des architectures et des règles de fonctionnement qui doivent être systématiquement vérifiées lors de toute activité. Par exemple dans le cas du réseau, il s'agit de vérifier que l'utilisateur qui y accède est effectivement autorisé à le faire, et que les interactions qu'il entretient avec le réseau sont conformes à ces règles.

Même si le fonctionnement de l'infrastructure est sécurisé, il est nécessaire en complément de sécuriser la donnée. La donnée est comprise ici dans un sens très général et peut inclure les traitements que cette donnée doit subir, algorithmes, codes, etc. Une démarche de cybersécurité en profondeur ne peut se reposer sur le fait que l'infrastructure soit sécurisée pour assurer la sécurisation des données, mais doit en complément déployer ses propres mécanismes de protection. Dans le cas du réseau, cela peut prendre la forme de tunnels chiffrés, sur différentes couches, comme des réseaux privés virtuels.

FORMALISATION DE LA CYBERSÉCURITÉ : LE *FRAMEWORK* NIST

Pour raisonner sur les problématiques de cybersécurité, nous nous référons au *framework* produit par le NIST³, et ici sa version 2.0. Les différents concepts du *framework* NIST sont résumés dans la Figure 3 page suivante.

³ National Institute of Standards and Technology.



Figure 3 : Concepts du framework NIST (Source : National Institute of Standards and Technology).

Tableau 1 : Les fonctions et catégories du framework NIST (Source : National Institute of Standards and Technology).

Fonction	Category	Identifier
Govern (GV)	Organizational content	GV.OC
	Risk management settings	GV.RM
	Cybersecurity supply chain risk management	GV.SC
	Roles, responsibilities and authorities	GV.RR
	Policies, processes and procedures	GV.PO
	Oversight	GV.OV
Identify (ID)	Asset management	ID.AM
	Risk assessment	ID.RA
	Improvement	ID.IM
Protect (PR)	Identity management, authentication and access control	PR.AA
	Awareness and training	PR.AT
	Data security	PR.DS
	Platform security	PR.PS
	Technology infrastructure resilience	PR.IR
Detect (DE)	Continuous monitoring	DE.CM
	Adverse event analysis	DE.AE
Respond (RS)	Incident management	RS.MA
	Incident analysis	RS.AN
	Incident response reporting and communication	RS.CO
	Incident mitigation	RS.MI
Recover (RC)	Incident recovery plan execution	RC.RP
	Incident recovery communication	RC.CO

Le *framework* NIST (<https://www.nist.gov/cyberframework>) est organisé en six fonctions, chacune découpée en plusieurs catégories indiquant les différents éléments à prendre en compte. Les cinq fonctions hors gouvernance (considérée comme transverse) relèvent d'une démarche d'analyse de risque, dans laquelle l'organisation commence par identifier ses biens et ses risques, puis met en place les mécanismes de protection nécessaires. Les risques non couverts sont ensuite traités par des mécanismes de détection, puis de réponse et de reprise si une attaque amène une compromission.

La gouvernance

La gouvernance de cybersécurité d'une organisation établit les différents éléments organisationnels qui vont piloter la politique de cybersécurité, tant sur le plan humain que sur le plan des règles et des process. C'est un point particulièrement critique pour les nouveaux réseaux, puisque le fournisseur d'infrastructure réseau et *cloud* peut héberger des services tiers, et doit inter-opérer avec d'autres fournisseurs d'infrastructure. Chaque fournisseur doit donc établir les règles d'interaction avec les organisations qu'il héberge et ses pairs, le déploiement des politiques de cybersécurité, et le traitement des incidents. Il doit également définir des mécanismes de confiance pour dialoguer avec eux et valider les informations échangées.

L'identification des risques

L'identification des risques repose sur l'identification des composants matériels et logiciels d'une part, sur la connaissance des vulnérabilités et des attaquants potentiels d'autre part. Cette identification des composants est plus difficile dans un contexte de réseaux du futur car ils sont par nature définis par le logiciel, et donc peuvent changer de surface d'attaque très rapidement. Certains composants, matériels ou logiciels, sont par ailleurs actifs dans le réseau mais ne sont pas la propriété de l'opérateur, ce qui accroît la difficulté de connaître leur fonction et leurs vulnérabilités.

La protection

La protection contre les attaques couvre de nombreux domaines, techniques et organisationnels. Ces mécanismes de protection forment la première barrière face aux attaquants, et ils sont les éléments clés de la mise en œuvre d'une politique de cybersécurité solide.

Les réseaux du futur vont cependant nécessiter des adaptations significatives de ces mécanismes classiques et bien maîtrisés. Leur taille et leur nombre d'utilisateurs vont rendre difficile le déploiement des outils d'identification, d'authentification et de contrôle d'accès, à cause de l'échelle, et également de la dynamique d'usage de ces réseaux lorsqu'il s'agira de fournir des services à la demande. L'ouverture de services à la demande implique une compétition pour l'accès à des ressources limitées, ce qui offre tout d'abord des possibilités d'attaque par déni de service. Plus significatif, la colocalisation de services peut rendre possible des fuites de données ou l'espionnage d'un client par un autre. Il sera donc nécessaire d'inventer de nouveaux mécanismes d'isolation, tant réseau que système, et d'exposer ces mécanismes aux clients pour leur permettre de piloter également leur propre politique de cybersécurité.

Détection

La détection des attaques s'est imposée comme un complément indispensable aux mécanismes de protection, qui malgré leur efficacité peuvent être submergés par des vagues importantes d'attaques, ou contournés par des attaquants très puissants. Il sera donc nécessaire de mettre en place dans les réseaux du futur des sondes permettant d'observer le comportement du réseau, et de l'analyser en temps réel pour détecter des modifications du comportement qui pourraient être symptomatique d'attaques.

En réponse aux besoins réglementaires, et pour optimiser la détection et la réponse, ces réseaux du futur devront être capables d'échanger des informations relatives aux vulnérabilités et aux incidents de sécurité. La mise en place de ce partage d'informations, par nature sensibles, devra également s'appuyer sur des mécanismes techniques et sur des relations de confiance entre opérateurs.

Réponse et remédiation

La réponse et la remédiation forment deux éléments distincts du *framework* NIST, qui sont cependant très proches. La réponse traite de l'incident dès qu'il est détecté, et a pour but de limiter l'impact de la compromission sur le fonctionnement du réseau. Cela implique des aspects techniques, comme la capacité à tracer les activités des attaquants pour identifier les composants compromis, et des aspects organisationnels, comme la communication de crise. Les infrastructures numériques à venir doivent être conçues en prenant en compte ce besoin de traçabilité, qui est souvent mal couvert dans les infrastructures actuelles.

À plus long terme, la remédiation a pour but d'éviter que la compromission se reproduise, et fait évoluer la configuration et le fonctionnement du réseau pour que ces attaques ne puissent plus réussir. Cela implique généralement des modifications dans la politique de cybersécurité.

Où en sommes-nous en France ?

Dans le cadre français, nous opérons les infrastructures numériques selon les réglementations européennes (RGPD, NIS et NIS2, DSA et DMA, et d'autres à venir). Nous participons à l'élaboration de ces règlements et nous sommes globalement actifs pour en assurer le bon déploiement sur le territoire national. À ce titre, l'Agence Nationale de Sécurité des Services d'Information (ANSSI) assume un *leadership* reconnu pour protéger les services de l'État, et peut intervenir dans des cas critiques.

En termes industriels, il existe un écosystème de grands groupes et de PME avec de fortes compétences en cybersécurité. Cependant, la plupart des produits déployés sont d'origine étrangère (américaine, israélienne, etc.). Un produit de cybersécurité ne peut être rentabilisé sur un marché national. L'émergence d'une industrie européenne de la cybersécurité nous semble absolument indispensable pour assurer l'indépendance technologique et la souveraineté numérique de chacun des États membres.

Conclusion : faire face aux nouvelles menaces

De nouvelles menaces apparaissent régulièrement, qui doivent être prises en compte. Plusieurs éléments ont cependant émergé ces dernières années, qui devront impérativement être pris en compte par les infrastructures numériques du futur :

- Les attaques sur la chaîne logistique logicielle : nous avons depuis quelques années ce type d'attaques qui touchent autant les logiciels fermés (exemple : SolarWinds) que les logiciels libres (exemple : LZ). Ces attaques ont des conséquences très importantes sur la maîtrise du logiciel inclus dans ces infrastructures, avec des conséquences majeures sur les privilèges que les attaquants obtiennent grâce aux chevaux de Troie inclus dans cette chaîne logistique logicielle. Il deviendra impératif de sécuriser cette chaîne logistique, pour assurer tant la disponibilité que l'intégrité du logiciel.
- Les attaques contre les modèles et algorithmes d'intelligence artificielle : si les attaques contre les outils d'IA ne sont pas nouvelles, elles ont un impact sensiblement plus grand pour des systèmes contrôlés par des IA. La complexité des algorithmes et la disponibilité des données et des matériels nécessaires à l'apprentissage peut

poser des problèmes de souveraineté, dans la capacité à se procurer lesdites données, matériels (GPU) et aussi les compétences nécessaires pour opérer les mécanismes d'apprentissage.

- La complexité des systèmes et des frontières : il devient de plus en plus difficile de comprendre le fonctionnement de systèmes numériques de grande taille, donc de définir leur politique de cybersécurité. Le nombre d'acteurs va également rendre nécessaire la délégation de responsabilités, ce qui engendre de nouveaux besoins d'audit et de contrôle.

Les infrastructures numériques doivent impérativement offrir à leurs utilisateurs des solutions de communication, de calcul et de stockage de confiance et être pérennes. Nous disposons de nombreux mécanismes pour protéger ces infrastructures, malgré le fait que les attaques cyber aient un impact significatif sur le fonctionnement de ces infrastructures. Les mécanismes de protection ont également un coût technologique et économique.

Le renforcement des mécanismes existants pour les adapter aux besoins des cas d'usage modernes (faible consommation d'énergie, faible latence, faible débit, etc.) des infrastructures critiques est absolument nécessaire pour faire émerger des infrastructures numériques de confiance, ou les aspects liés à la cybersécurité (contrôle d'accès et d'usage, gestion des identités et des accès, audit et contrôle) sont compris et acceptés par tous les utilisateurs de ces infrastructures. Le lien entre cybersécurité et sûreté de fonctionnement, longtemps considéré comme important mais difficile à mettre en œuvre, doit également être renforcé pour assurer, dans le cas des infrastructures critiques, à la fois la sécurité des biens et des personnes.

Finalement, comme pour beaucoup de sujets touchant au numérique, conserver sa souveraineté demande à la fois des capacités technologiques pour développer ou déployer des outils, mais aussi les capacités humaines pour inventer de nouveaux outils ou pour opérer de manière adéquate des outils sensibles dans des environnements complexes, les conséquences d'une erreur pouvant être majeures.

Les composants pour les infrastructures numériques

Par Dimitri KTÉNAS et le Dr Denis DUTOIT

Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA)

Dans cet article, nous nous intéressons plus particulièrement aux composants constitutifs des infrastructures numériques, en premier lieu sous l'angle des infrastructures de calcul, puis sous l'angle des infrastructures de communication. Les composants clés de l'infrastructure de calcul des systèmes numériques incluent les capteurs intelligents, les processeurs d'application, et tout xPU fournissant une capacité de calcul. Ces composants créent une infrastructure de traitement numérique hétérogène, flexible et évolutive, capable de répondre aux besoins croissants du calcul.

D'un autre côté, le domaine des télécommunications est devenu la clé de voûte de l'économie des pays occidentaux. Il apparaît que la 6G, en cours de définition, va nécessiter des composants dédiés nouveaux du fait notamment de la montée en fréquence qui risque de rendre inopérant les composants actuels en silicium. C'est donc une véritable rupture qui se prépare dans ce domaine de la microélectronique.

INTRODUCTION

L'évolution du monde numérique a toujours suivi celle des besoins de l'utilisateur, de la facilité d'utilisation, des services offerts. Tout a commencé avec le besoin d'exécuter des calculs toujours plus nombreux et complexes, puis de rapprocher la source d'information au plus près de l'utilisateur. En parallèle sont apparus internet et le *smartphone*, lequel est aujourd'hui le moteur des avancées les plus importantes dans les technologies et les composants, y compris dans les nœuds technologiques très avancés.

Les évolutions liées à la transformation numérique sont supportées par des technologies sous-jacentes destinées à un déploiement massif telles que l'internet des objets, l'intelligence artificielle, la cybersécurité, le jumeau numérique. Avec nombre de défis à relever : qualité, sûreté, sécurité des services, gestion et traitement des flux massifs de données, miniaturisation de composants toujours plus performants, le tout avec des contraintes fortes de réduction de la consommation énergétique. Dans la suite de ce chapitre, nous nous intéressons plus particulièrement aux composants constitutifs des infrastructures numériques, en premier lieu sous l'angle des infrastructures de calcul, puis sous l'angle des infrastructures de communication. L'objectif poursuivi est de présenter les challenges à relever pour ces composants mais aussi de donner un aperçu de ce qu'ils pourraient être dans le futur.

INFRASTRUCTURE DE CALCUL DES SYSTÈMES NUMÉRIQUES

L'augmentation massive des données liées aux dispositifs IoT (*Internet-of-Things*), à la diffusion de contenus multimédias ainsi qu'à l'Intelligence Artificielle engendre une transformation majeure de l'infrastructure de calcul des systèmes numériques. Cette

évolution est bien détaillée dans les différentes feuilles de route du projet européen HiPEAC (<https://www.hipeac.net/#/>) depuis 2004. La Figure 1 résume cette évolution en trois phases :

- *Cloud computing* : initialement, le calcul et le stockage étaient principalement centralisés dans les centres de données traditionnels. Les capteurs du monde physique remontaient les données brutes vers ces centres pour y être traitées.
- Systèmes cyber-physiques intelligents : avec l'augmentation des dispositifs IoT et une volonté de gagner en efficacité énergétique, l'architecture a évolué pour inclure des composants de traitement numérique plus proches de la source de données. Ainsi, les microcontrôleurs et les capteurs IoT sont devenus plus sophistiqués, capables de prétraiter les données avant de les envoyer à des passerelles (*gateway*) ou des serveurs à la périphérie (*edge computing*). Ces serveurs ont gagné en puissance de calcul et sont maintenant capables de traiter des applications d'inférence en intelligence artificielle, réduisant ainsi le besoin de transfert de données volumineuses vers le *cloud* (F. Palumbo *et al.*, 2023).
- Systèmes distribués coopératifs : les besoins applicatifs à la périphérie vont encore évoluer vers plus de puissance de calcul, plus de données à traiter, le tout avec une faible latence. L'exemple le plus probant étant l'intelligence artificielle générative. Les capacités de calcul à la périphérie ne seront plus suffisantes, notamment pour l'apprentissage incrémental de ces modèles. C'est pourquoi une nouvelle couche de petits serveurs distribués et coopérant entre eux, situés plus près des utilisateurs finaux, va se créer pour offrir une puissance de calcul intermédiaire. Ainsi, les ressources de stockage et de traitement seront réparties entre le *cloud* centralisé dans les centres de données, le *cloud* distribué, les serveurs à la périphérie et enfin les capacités de calcul des capteurs pour former le continuum de calcul ou continuum informatique. Ce modèle permet d'optimiser la gestion des données et les processus de calcul en fonction de leur emplacement et de leurs exigences spécifiques.

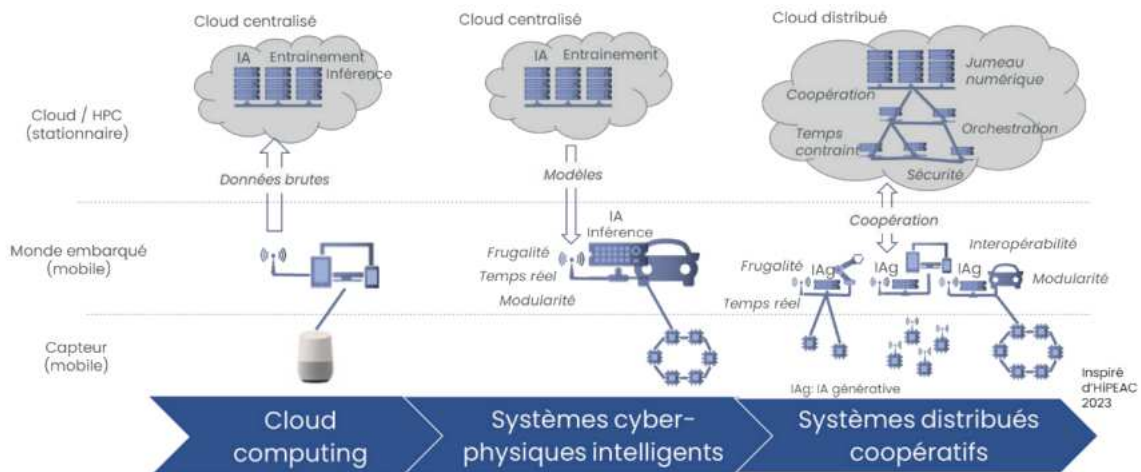


Figure 1 : Évolution du continuum informatique (Source : HiPEAC 2023).

Les composants clés de ce continuum incluent :

- les capteurs intelligents pour la couche de calcul proche du monde physique et ainsi collecter et prétraiter les données localement pour réduire la consommation de bande passante et d'énergie ;
- les processeurs d'application des téléphones mobiles ou encore les processeurs dits de fusion pour l'électronique de l'automobile offrent une puissance de calcul intermé-

diaire au niveau de la périphérie pour fusionner et traiter des données provenant des capteurs ;

- enfin, les CPU, GPU, TPU ou tout autre xPU fournissent une capacité de calcul à grande échelle dans le *cloud* centralisé et à plus petite échelle dans le *cloud* distribué plus proche des utilisateurs.

Ensemble, ces composants créent une infrastructure de traitement numérique hétérogène, flexible et évolutive, capable de répondre aux besoins croissants du calcul. L'évolution des composants dépend de leur position dans le continuum de calcul, comme expliqué ci-après.

Pour les composants proches du monde physique (capteurs), le principal challenge réside dans l'atteinte de l'ultra-basse consommation. Les composants devront ainsi être de plus en plus autonomes au niveau énergétique, avec des architectures en rupture incluant le calcul directement au sein des capteurs *via* une intégration multicouches 3D (*i.e.* un assemblage vertical de puces) hétérogène. En d'autres termes, l'architecture du composant incorporera une couche capteur, une couche pré-traitement, une couche mémoire et une couche de traitement de l'Intelligence Artificielle (IA). Ce type de solution est déjà utilisée, par exemple dans le cas des capteurs d'images pour la téléphonie mobile. La France est bien située dans le domaine de l'intégration 3D, que ce soit au niveau de la recherche avec l'IRT Nanoélectronique ou au niveau industriel avec les imageurs de ST Microelectronics ou encore de la société Prophesee, dans ce que l'on appelle la vallée des capteurs d'images aux alentours de Grenoble. Dans un autre registre, une société comme Greenwaves propose des processeurs d'application très basse consommation pour les objets connectés.

Pour les composants du *cloud*, le challenge a basculé ces dernières années sur le traitement de l'inférence et de l'apprentissage des modèles d'IA génératives. En 2022, un modèle d'IA générative (par exemple GPT-3) comporte déjà 200 milliards de paramètres alors que la mémoire de travail d'un GPU approche seulement les quelques dizaines de milliards d'octets. Et cet écart va en s'accroissant. La mémoire et son utilisation efficace en énergie deviennent désormais le principal challenge de ce segment du continuum. Les mémoires de type "High Bandwidth Memory" utilisées pour ces applications d'intelligence artificielle se feront de plus en plus denses avec l'assemblage en vertical de plus de 12 couches de puces mémoires et une bande passante avec le GPU de plus en plus élevée comme les futures HBM4. Les GPU deviennent les seuls composants capables de traiter les modèles d'IA générative dans le *cloud* avec une domination de NVIDIA et son dernier Blackwell, fabriqué en 4 nm et qui contient plus de 200 milliards de transistors. AMD avec son MI300 se positionne comme le principal concurrent de NVIDIA. Pour encore augmenter les capacités de calcul, une société californienne récente, Cerebras, utilise une tranche de silicium complète pour son processeur (wafer-level computing). C'est donc la course au gigantisme à laquelle la France n'est pas du tout présente. Pour le *cloud*, la France n'apparaît en effet que pour les CPU avec l'Initiative de Processeur Européen (EPI) d'EuroHPC et la conception du processeur Rhea de la toute récente société SiPearl. Une autre évolution majeure du *cloud* réside dans l'intégration de la photonique dans les communications entre les racks des serveurs. Cette communication optique, arrive désormais entre les boîtiers des composants avec des circuits d'entrées/sorties électro-optiques pour gagner en latence de transmission et ainsi accroître encore la capacité mémoire et de traitement d'un nœud de calcul. Enfin devrait arriver d'ici une dizaine d'années le calcul analogique optique où les multiplications-accumulations ne seront plus numériques mais optiques. Cela permettra de franchir plusieurs ordres de grandeurs en capacité de calcul et en efficacité énergétique. Mais de nombreux challenges restent à relever dans ce domaine comme l'énergie nécessaire pour les conversions électro-optiques inévitables (les calculs non linéaires ne peuvent pas se faire en optique) ou la taille des filtres optiques mis en œuvre pour le calcul.

Pour l'embarqué, les performances de calcul et les capacités de mémoire sont d'un à deux ordres de grandeurs moindres que pour le *cloud*. Pour garder une bonne efficacité énergétique, des architectures de rupture seront tout de même nécessaires comme le calcul proche/dans la mémoire. Beaucoup de société *fabless* se créent autour de ce segment de l'Intelligence Artificielle en périphérie (embarquée) et aussi en France avec Kalray, VSORA, Menta et d'autres encore. ST Microelectronics se positionne également sur ce segment avec ses microcontrôleurs pour le marché de l'automobile. C'est dans l'embarqué, pour l'inférence en IA que le calcul analogique a retrouvé un regain d'intérêt avec plus particulièrement les circuits neuromorphiques qui tentent d'émuler un réseau de neurones en remplaçant les multiplications-accumulations numériques par un maillage de résistances où les courants s'additionnent. Le challenge de ces architectures analogiques est d'avoir un réseau de résistances suffisamment dense pour émuler un réseau complet. Pour cela, les nouvelles mémoires résistives non-volatiles sont prometteuses.

Si l'on se projette plus avant, à l'horizon 2040, on peut raisonnablement anticiper une interconnexion accrue des systèmes informatiques, associée à l'intégration de nouvelles technologies et architectures dans les composants. Cette évolution augmentera les capacités de calcul et l'efficacité énergétique, tout en permettant une granularité plus fine des systèmes distribués. Cependant, cette évolution n'est pas sans risque. Les travaux de C. Weir *et al.* (2024) identifient quelques impacts négatifs de cette évolution comme la difficulté de distinguer la réalité de la fiction due à la prolifération de la désinformation générée par l'IA ou encore la difficulté de comprendre le comportement des systèmes et ainsi de différencier une attaque intentionnelle d'un incident mineur. Pour atténuer ces risques, les auteurs préconisent quelques recommandations comme la « responsabilité ambiante » ou plus précisément la capacité des systèmes à se vérifier eux-mêmes pour garantir la sécurité ou encore l'investissement dans des méthodes de développement responsable avec de meilleures pratiques en ingénierie logicielle pour les systèmes d'IA.

INFRASTRUCTURE DE COMMUNICATION DES SYSTÈMES NUMÉRIQUES

Le domaine des télécommunications, et plus particulièrement des communications sans-fil, est aujourd'hui devenu la clé de voute de l'économie des pays occidentaux. Depuis l'avènement des *smartphones*, c'est également un domaine structurant pour la microélectronique. Le déploiement de la 5G met en évidence l'importance pour notre souveraineté de maîtriser les technologies de communication. Les batailles technologiques et une partie des tensions géopolitiques d'aujourd'hui sont nées autour des technologies de communication (3G, 4G, 5G et demain 6G). La France garde un historique fort sur ce domaine malgré l'évolution des grands acteurs (acquisition d'Alcatel par Nokia, opérateur historique se tournant vers le service) et l'Europe héberge toujours des grands industriels (Nokia, Ericsson) du secteur (principalement l'infrastructure). Tout comme les précédentes générations de systèmes de télécommunication, le déploiement de la 5G est rendu possible grâce aux avancées de la microélectronique (Loi de Moore), qui permet un traitement de l'information plus efficace énergétiquement. Les gains visibles pour les utilisateurs sont une connectivité plus facile amenée par des débits plus élevés et des temps de latence plus faibles. À un moindre degré, une meilleure couverture est également attendue. Aussi, ces technologies de communication ont largement dopé le marché des semi-conducteurs.

Des travaux sont d'ores et déjà en cours pour définir le périmètre et les spécifications de la 6G. Cette génération devrait arriver sur le marché à l'horizon 2030. D'emblée, il apparaît que cette génération va nécessiter des composants dédiés nouveaux du fait notamment de la montée en fréquence qui risque de rendre inopérant les composants actuels en silicium. C'est une véritable rupture qui se prépare donc dès maintenant dans ce domaine de la microélectronique avec l'utilisation de matériaux nouveaux (introduc-

tion de matériaux III-V) plus adaptés aux fréquences hautes. Plus précisément, la 6G va s'accompagner d'une montée importante en fréquence des bandes de communication, permettant plus de bande passante, de débit et une latence réduite, depuis les bandes millimétriques (> 12 GHz) vers les bandes sub-THz. Cette montée en fréquence changera drastiquement la propagation des ondes et *in fine* les réseaux puisque même en considérant des communications directives, par opposition aux antennes omnidirectionnelles ou à la segmentation de l'espace actuels, les portées seront plus faibles à ces fréquences avec donc beaucoup plus de stations de base envisagées.

En parallèle une autre révolution s'opère dans les infrastructures télécoms avec le déploiement de nouveaux Radio Access Networks (voir la Figure 2) : entre les « têtes RF » / RRU – Remote Radio Unit – délivrant et recevant les signaux, et le cœur de réseaux fournissant les données / CN – Core Network, l'architecture est devenue entièrement reconfigurable et programmable (virtualization-RAN), répartie entre Distributed Unit – DU – très nombreuses et Centralized Unit – CU – regroupant les précédentes.

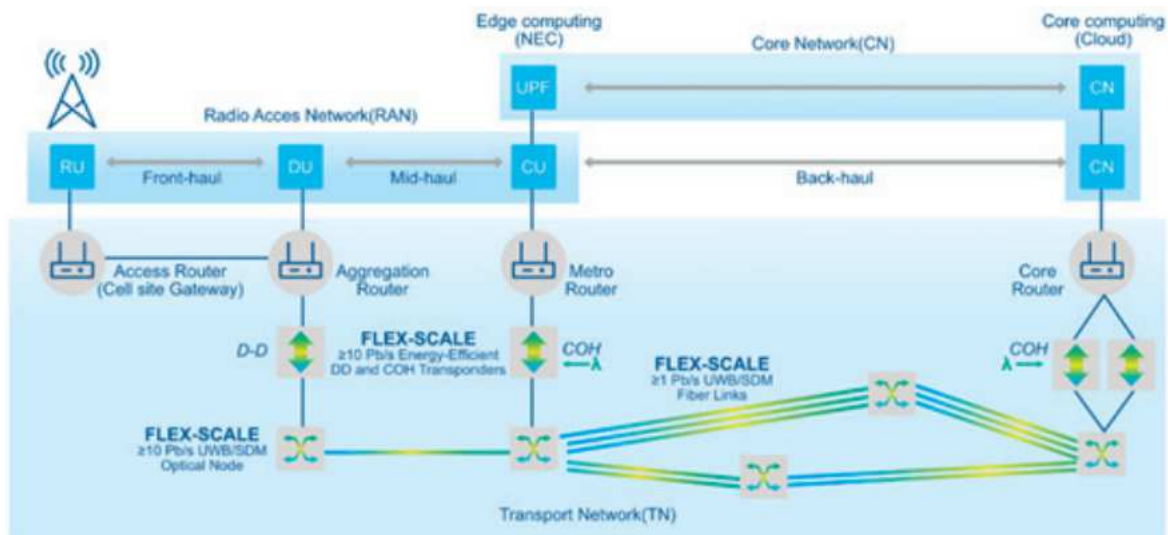


Figure 2 : Réseau de bout-en-bout vu par l'ETSI pour le B5G ou 5.5G (Source : Effenberger, 2022).

Cette reconfigurabilité et programmabilité implique que les acteurs fournisseurs de ces solutions changent et que seront favorisés les fournisseurs de solutions matérielles et logicielles les plus performantes. Il sera alors plus facile d'insérer ces RRU fonctionnant en bandes hautes au sein d'un réseau offrant plus d'opportunités de découpage et de virtualisation, avec à chaque niveau la possibilité de mise en œuvre d'outils logiciels qui favoriseraient les acteurs américains ou japonais, dans un premier temps, déjà bien connus au détriment des fournisseurs d'infrastructures dédiées télécom tels que Nokia et Ericsson. La nouvelle architecture portée par l'Open RAN nécessitera d'avoir des processeurs au niveau des DU et des CU. Une première indication semble favoriser un acteur tel que Nvidia qui est *leader* dans le domaine des processeurs graphiques, très bien adaptés aux besoins des réseaux, et conçus dans des technologies FinFet très fines. En Europe, un acteur montant comme Kalray pourrait jouer un rôle essentiel dans l'avènement d'une solution souveraine, utilisant des technologies continentales.

Concernant la fourniture de composants pour les télécoms, les acteurs européens n'ont qu'une petite part face à des concurrents tels que Qualcomm, MediaTek, NVidia, Apple, Intel/Altera, AMD/Xilinx... qui sont au premier plan pour les terminaux, mais également pour les réseaux. Les acteurs européens majeurs – STMicroelectronics, NXP, Infineon –

sont en revanche *leaders* pour les composants en lien avec les marchés verticaux associés aux applications industrielles.

De leurs côtés, STMicroelectronics et GF proposent des technologies adéquates pour des composants RF mais également des processeurs adaptés à certaines applications : FD-SOI 28 nm et 22 nm, avec évolution vers le 18 nm puis le 10 nm, FinFet 12 nm pour GF, BiCMOS 55 nm pour ST. Ces technologies fines proposées ont déjà démontré leurs intérêts pour des applications RF, c'est-à-dire au niveau des RRU dans les réseaux, pour la connectivité *wifi* et pour les circuits bas coût de l'IoT. À l'inverse, ils ne disposent pas de finesse de gravure suffisante – 5 nm et en dessous – pour entrer dans les terminaux mobiles. Les enjeux pour les acteurs français et européens résident dans leur capacité à répondre au besoin en approvisionnement de composants sur des parties bien déterminées du réseau.

Dans les RRU, les têtes RF seront de différents types, en fonction de la fréquence et des performances demandées. Des industriels européens tels que STMicroelectronics peuvent proposer des technologies hybrides basées sur le B9MW (130 nm BiCMOS) et le B55X (55 nm BiCMOS) et co-intégrant des technologies de type GaN pour améliorer les performances en efficacité énergétique en sub-7 GHz et bandes K, et de type InP pour la montée en fréquence pour les bandes mmW 60 GHz, 140 GHz voire 300 GHz (Calvanese *et al.*, 2021). De même, pour les réseaux, l'Europe a une carte à jouer dans le domaine de l'optique et notamment de la photonique sur silicium. Les réseaux télécoms vont être massivement déployés autour de réseaux de fibres optiques dont les composants de tête – lasers et diodes – co-intègrent des technologies II-VI pour l'optique et CMOS/BiCMOS pour les parties électroniques. Ce type de co-intégration correspond parfaitement au modèle des fonderies européennes, qui peuvent fournir les technologies de base et approvisionner celles manquantes pour assurer l'assemblage.

CONCLUSION

L'émergence de cœurs de réseaux virtualisés va apporter de la flexibilité et permettre la connexion de nombreuses modalités de communication sans fil autour d'un traitement numérique très puissant et flexible. Une complexité globale accrue, et en corollaire, une maîtrise difficile de l'ensemble des données échangées font apparaître une difficulté en matière de souveraineté et la possibilité de captation d'informations. La tendance actuelle de rachat par les Gafam, ou d'alliances stratégiques dans le monde des Telecom, ne laisse aucun doute sur l'importance de la maîtrise de ce cœur de réseau et de la rupture à accomplir en Europe pour rattraper ces acteurs mondiaux américains. Par-dessus cette maîtrise du monde numérique, il faut aussi considérer l'importance de la gradation des traitements opérés : du *mobile edge computing* proche de la source d'information au *cloud* générique et hyper-concentré, différents niveaux vont nécessiter d'être développés et maîtrisés.

Autour de ce réseau central, la variété des modalités RF va multiplier les systèmes radio déployés : plus de composants, de matériel et d'objets technologiques disséminés. Il conviendra donc d'être attentif à la neutralité écologique de ces systèmes, à l'acceptation sociétale et en regard des besoins accrus en composants, être assuré de la viabilité de la filière microélectronique européenne pour réduire la dépendance actuelle à la Chine, à Taïwan et aux États-Unis. Des telles technologies existent déjà pour répondre à une partie des besoins, notamment les technologies à base de SOI (FD-SOI ou RF-SOI), mais l'évolution rapide des communications sans-fil, et l'accroissement de connectivité dans les applications va nécessiter de considérer l'élargissement des solutions actuelles vers plus de diversité, en particulier vers des nœuds technologiques plus avancés.

BIBLIOGRAPHIE

PALUMBO F., LAZCANO R. & MADRONAL D. (2023), “Towards a living dimension: The future of cyberphysical systems”, in DURANTON M. *et al.*, editors, *HiPEAC Vision 2023*, pp. 44-53.

WEIR C., DYSON A., JOGUNOLA O., DENNIS L. & PAXTON-FEAR K. (2024), “Inter-linked computing in 2040: Safety, truth, ownership, and accountability”, *Computer*, vol. 57, n°1, pp. 59-68.

EFFENBERGER F.J. (2022), “Fixed 5th generation advanced and beyond”, 1st edition – September, ETSI White Paper N°#50, ISBN N°979108262071.

CALVANESE STRINATI E., BELOT D., FALEMPIN F. & DORE J.B. (2021), “Towards 6G: from new hardware design to wireless semantic and goal-oriented communication paradigms”, ESSCIRC 2021, 6-20 September 2021.

Apport de l'infrastructure de partage de données Agdatahub dans le secteur agri-agro

Par Sébastien PICARDAT
Directeur général d'Agdatahub

Ultra-connectée, la Ferme France génère énormément de données. L'infrastructure de partage de données Agdatahub est le tiers de confiance du secteur agricole et agroalimentaire *via* lequel les données peuvent circuler en conformité avec le cadre réglementaire européen. En effet, la plateforme française est un outil technologique et opérationnel qui répond aux objectifs de durabilité, d'innovation et de souveraineté poursuivis par les acteurs des filières. En tant qu'intermédiaire de données, statut inscrit dans la réglementation européenne et la loi française, Agdatahub simplifie et sécurise la circulation des données agricoles et traite les cas d'usage jugés prioritaires comme la génétique animale, la réduction du carbone, l'affichage environnemental, etc.

LA FERME FRANCE, GRANDE CONSOMMATRICE DE NOUVELLES TECHNOLOGIES

En 2024, les exploitations agricoles ne produisent plus seulement de quoi nous nourrir. Elles produisent aussi une masse énorme de données... Les agriculteurs ont de tout temps été des technophiles. Ils ont ainsi été les premiers à adopter le minitel dans leur quotidien professionnel. Dans les années 1980-1990, la profession avait ainsi recours à de nombreux outils « en ligne » afin de garder à l'œil les prévisions météorologiques ou encore le cours des matières premières.

Depuis qu'internet a remplacé le minitel (France Télécom a définitivement interrompu le service en 2012) et que les solutions numériques se sont développées pour tous les types de culture ou d'élevage, la Ferme France et ses 824 000 agriculteurs sont très connectés¹ : ils sont 81 % à utiliser quotidiennement Internet dans le cadre de leur profession. Ils sont même 85 % à mobiliser les applications *smartphone* et les réseaux sociaux pour améliorer et partager leurs activités, quand les trois-quarts d'entre eux possèdent au moins un outil robotisé et que la moitié d'entre eux utilisent les systèmes GPS pour améliorer la précision de leurs travaux.

L'adhésion rapide à ces nouvelles technologies par le secteur agricole s'explique par leurs apports : un meilleur rendement des fermes, une plus grande anticipation des tâches, une surveillance facilitée des cultures ou des bêtes, etc. Ainsi, la vétérinaire Annick Valentin-Smith souligne-t-elle, que, dès 2010, l'arrivée « de capteurs fiables et précis dont on a pu doter les animaux en particulier les bovins – en collier ou en boucle d'oreille entre autres – a permis de capter en continu une très grande quantité de données, de les transmettre,

¹ Chiffres issus du dossier de presse Agriculture et Innovation : Lancement de la French AgriTech, 30 août 2021.

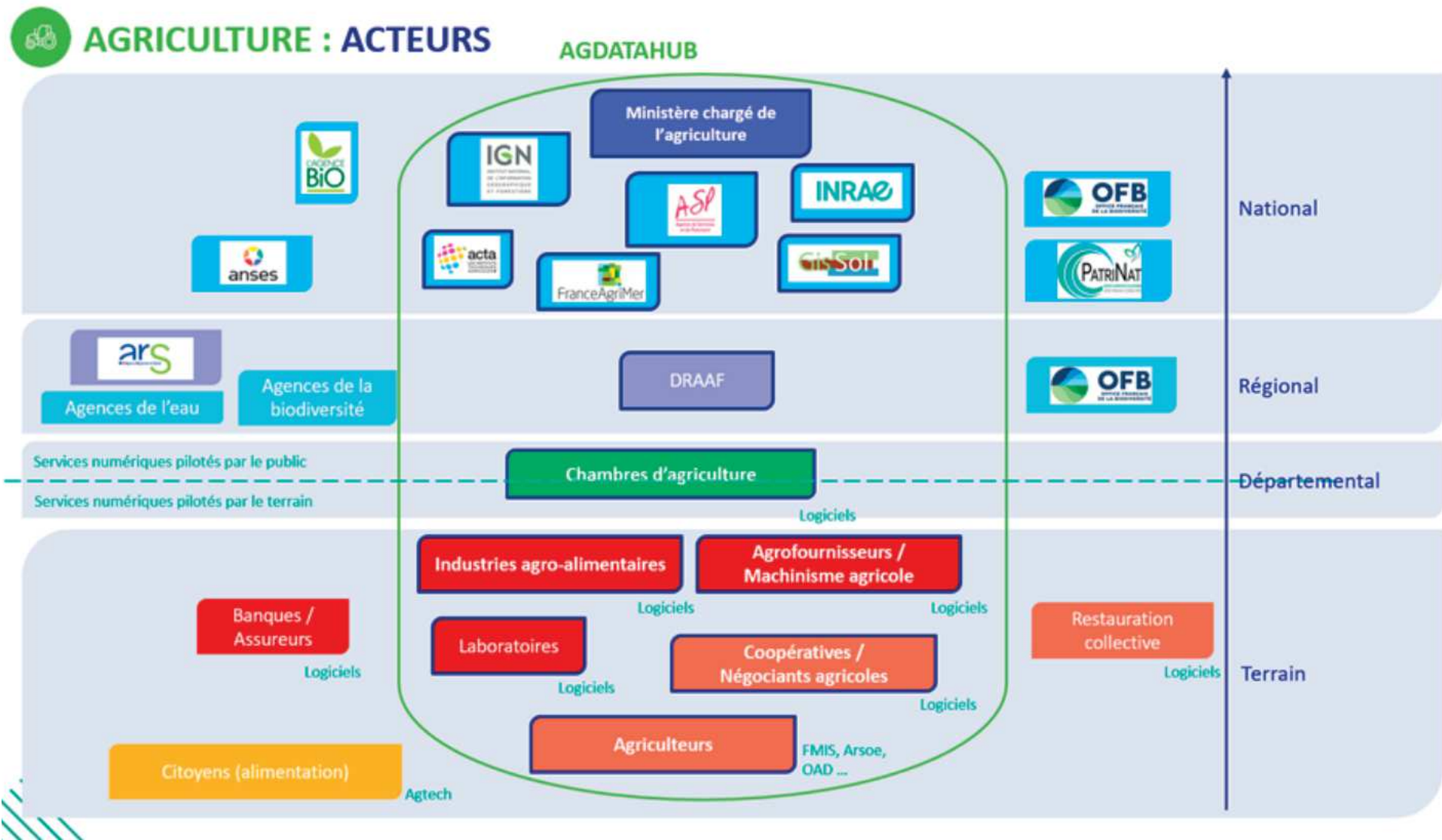


Figure 1 : Cartographie des acteurs dans le secteur agricole
 (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

de les analyser et de les exploiter² », ouvrant ainsi la voie aux premiers systèmes de télésurveillance, d'alerte et d'aide à la décision dédiés aux éleveurs.

UN PORTAIL POUR « UNE AGRICULTURE FRANÇAISE COMPÉTITIVE, OUVERTE ET SOUVERAINE »

L'essor de ces technologies dans les fermes et les élevages français s'accompagne comme on le sait de la production massive de données et c'est ainsi que la Ferme France est précocement entrée dans l'ère du *big data*. Voilà pourquoi, en 2015, le plan Innovation Agriculture 2025³ a été lancé par Stéphane Le Foll (alors ministre de l'Agriculture), Stéphane Mandon (secrétaire d'État à l'Enseignement supérieur et à la Recherche) et Axelle Lemaire (secrétaire d'État au Numérique). Le plan d'action qui en a découlé prévoyait notamment un programme de recherche sur « Technologies robotiques et Numérique pour l'agriculture » et « de façon complémentaire et transversale aux programmes les ministres ont souhaité la mise en place (...) d'un portail de données et services numériques pour une agriculture française compétitive, ouverte et souveraine »⁴.

Il s'agit là des prémices de ce que nous connaissons aujourd'hui comme la plateforme Exchange (ex-API-Agro) d'Agdatahub. L'objectif principal de ce portail, qui n'était alors qu'à l'état d'idée, était de « faciliter l'accès à des données et accélérer les processus de développement de connaissances, de modèles, et *in fine* de services à l'agriculture »⁵.

Déjà à l'œuvre pour le plan Innovation Agriculture 2025, Jean-Marc Bournigal, alors président de l'Irstea, a ainsi remis, en 2017, un rapport⁶ qui a posé les bases de ce qui allait être la première infrastructure française de partage de données agricoles. M. Bournigal y évoque aussi les enjeux (plus que jamais d'actualité) auxquels répond l'infrastructure Agdatahub : la durabilité, l'innovation, la souveraineté. L'intégration d'Agdatahub dans le rapport sur les infrastructures de partage de données, dont une synthèse est donnée dans le présent numéro, parachève cette reconnaissance institutionnelle.

« Comme dans les autres secteurs économiques, il y a en agriculture un enjeu fort à collecter et à traiter des données en vue d'en tirer des informations inédites et des modèles innovants sur la base desquels développer des services nouveaux. Cette opportunité n'a pas échappé non seulement aux géants de l'agrofourmiture qui cherchent à intégrer l'ensemble de la chaîne, de la collecte de données à la vente de produits accompagnés de services, mais aussi aux GAFAM (aujourd'hui, les GAFAM, NDLR) qui commencent à s'intéresser à l'agriculture, nouveau domaine d'action après le commerce, la banque et la santé. La concentration de telles ressources dans les mains de quelques acteurs génère des inquiétudes quant à l'indépendance des agriculteurs, voire la souveraineté alimentaire.⁷ »

² L'Intelligence Artificielle au service d'un élevage de précision, agdatahub.eu.

³ Communiqué de presse, octobre 2015, Agriculture – Innovation 2025 : des orientations pour une agriculture innovante et durable.

⁴ *Ibid.*

⁵ Communiqué du 10 janvier 2017, <https://agriculture.gouv.fr/un-portail-de-donnees-pour-linnovation-en-agriculture-la-synthese-du-rapport>

⁶ *Ibid.*

⁷ Rapport Bournigal : AgGate - Portail de données pour l'innovation en agriculture, p. 9.

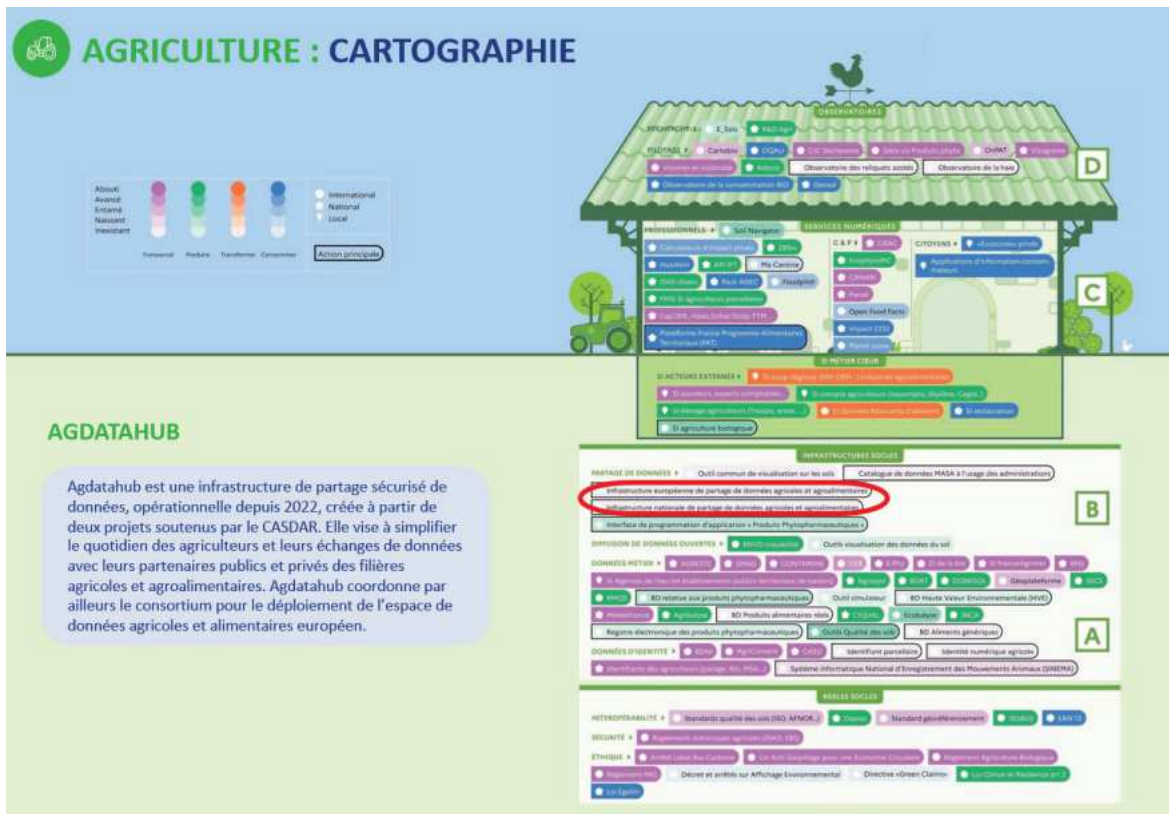


Figure 2 : Cartographie des composants *data* pour le secteur agricole (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

LE TIERS DE CONFIANCE DU SECTEUR AGRICOLE ET AGROALIMENTAIRE

Fruit des travaux de recherche API-Agro et Multip@ass soutenus par le Casdar (phase exploratoire de 2017), l'infrastructure de partage d'Agdatahub est opérationnelle depuis 2022. Elle est aujourd'hui dans une phase de passage à l'échelle. À sa solution initiale d'échanges de données (la plateforme repose sur la technologie Data Exchange Platform du Français Dawex, nommé Tech Pioneer en 2020 par le Forum Économique Mondial de Davos), sont venues s'ajouter des briques technologiques reposant sur la première identité numérique agricole (en co-innovation avec Orange Business et IN Groupe) afin de sécuriser les échanges de données et de noter et prendre en compte les autorisations des exploitations agricoles (émises par leurs gérants) à la circulation de leurs données, en conformité avec les évolutions réglementaires européennes telles que le Data Governance Act (DGA)⁸.

Ce texte européen institutionnalise le statut d'intermédiaire de données, véritable tiers de confiance mettant à disposition un cadre sécurisé pour simplifier et fluidifier la circulation des données entre détenteurs et utilisateurs. C'est la raison d'être de l'entreprise Agdatahub qui s'est d'ailleurs enregistrée au registre européen des intermédiaires de données depuis juin 2024.

⁸ <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>

FÉDÉRER DES OFFRES DE DONNÉES DÉCENTRALISÉES ET HÉTÉROGÈNES

Aujourd'hui, une exploitation agricole française peut abriter une trentaine de sources de données⁹ : robots de traite, GPS, télésurveillance des troupeaux, capteurs météo, sondes pédologiques, tracteur, moissonneuse-batteuse, pulvérisateur... Dire que les données agricoles sont issues de sources hétérogènes est un euphémisme ! Il s'agit donc de données décentralisées car elles sont stockées chez les différents partenaires des exploitations. Leur exposition au sein du catalogue centralisé de la plateforme Exchange d'Agdatahub (dans le respect des autorisations des exploitations et de manière publique ou en restreignant la visibilité), par fichier ou par API, donnent lieu à des échanges sécurisés et transparents entre partenaires identifiés et selon des critères fixés au cas par cas, librement, par ceux qui choisissent d'exposer leurs jeux de données.

L'interopérabilité au niveau européen est assurée par la fédération des catalogues des infrastructures et la fédération des identités numériques grâce à l'implémentation des standards de facto Gaia-X. En effet, Agdatahub est un des *lighthouse projects* de l'association Gaia-X et participe aux travaux de l'association, en plus d'avoir coordonné le consortium Agridataspace¹⁰ qui a livré la feuille de route du futur *data space* européen dédié à l'agriculture. À noter qu'Agdatahub a candidaté au nouvel appel à projet relatif à son déploiement, dans un nouveau consortium de 35 partenaires issus de 20 États membres.

LES DONNÉES POUR SUIVRE LES PERFORMANCES AGRO-ENVIRONNEMENTALES ET COMPTABLES DES EXPLOITATIONS

Aujourd'hui, l'infrastructure de partage de données Agdatahub répond à plusieurs cas d'usage émanant du secteur agricole ou agroalimentaire et jugés prioritaires par l'État, selon des objectifs fixés par France Nation Verte.

Par exemple, c'est le cas du suivi des performances agro-environnementales et comptables des exploitations. Pour toucher leurs aides de la PAC (Politique Agricole Commune), les exploitants agricoles doivent respecter plusieurs critères. Il leur faut alors renseigner leurs informations dans de multiples outils : dans leur système d'information de gestion agricole, dans Télépac (téléservices des aides de la PAC), dans leurs outils d'aide à la décision (OAD), etc. Outre les risques d'erreur et le temps perdu lors de ces multiples re-saisies d'informations, les exploitants encourent le risque de perdre la maîtrise des usages de leurs données. Compte tenu de la multiplicité des éléments à contrôler, les phases de contrôle sur leurs exploitations sont des moments qu'ils estiment stressants.

Le contrôleur PAC doit pouvoir vérifier simplement et sereinement que l'exploitation répond aux critères de conditionnalité, grâce à l'accès aux données requises pour le contrôle. Les acteurs du conseil agricole (les chambres d'agriculture, les coopératives, les négociants) doivent pouvoir accéder aux données des agriculteurs dans le respect de leurs consentements afin de les accompagner en amont des contrôles, si leurs clients le souhaitent.

Les acteurs de la recherche (comme l'Inrae et les instituts techniques agricoles) doivent, eux, accéder aux données agro-environnementales (issues des satellites, des capteurs

⁹ Sébastien Windsor, président de Chambres d'agriculture France, <https://reseaux.orange.fr/actualites/5g-et-numerique-secteur-agriculture.html>

¹⁰ <https://agridataspace-csa.eu/>

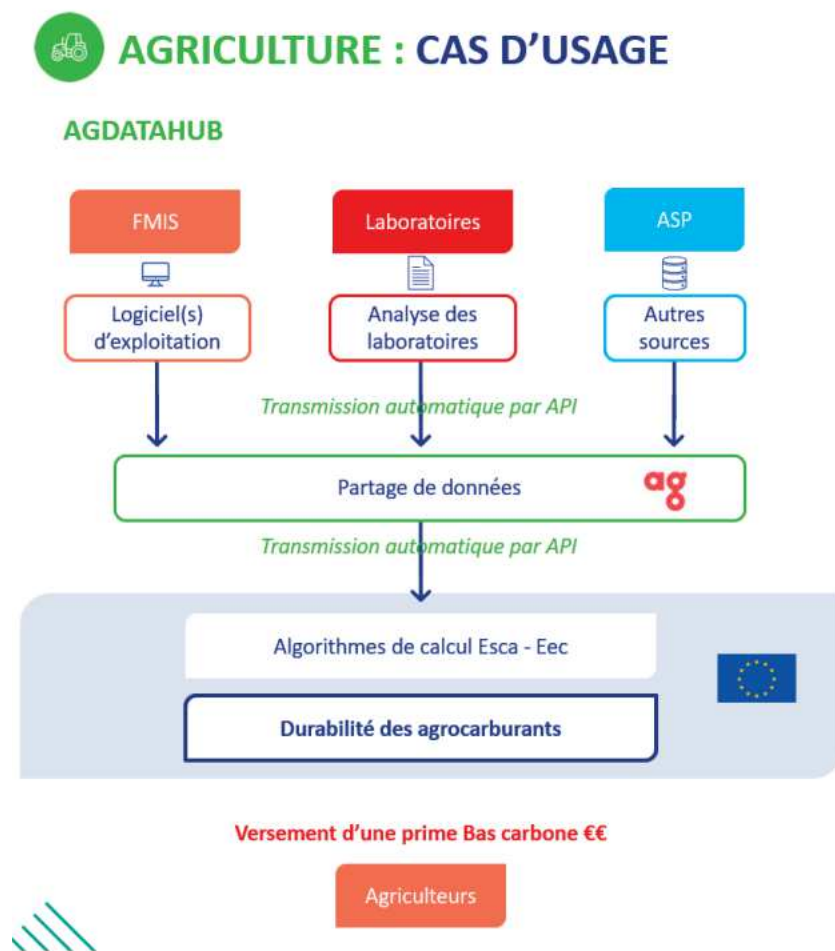


Figure 3 : Cas d'usage (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

dans le sol ou des capteurs météo) et comptables afin de calculer la performance agroenvironnementale (nutriments, émissions de gaz à effet de serre, biodiversité et santé des sols) et les performances économiques (taille de l'exploitation, revenu net, coût de production et travail non rémunéré).

Le ministère de l'Agriculture et de la Souveraineté alimentaire doit accéder aux données des agriculteurs et des instituts de recherche dans le respect des règles de confidentialité statistique afin de publier des études statistiquement représentatives et éclairer les décideurs publics.

IDENTIFICATION ET DONNÉES GÉNÉTIQUES POUR L'AMÉLIORATION DE LA PRODUCTION ANIMALE

Les acteurs de l'identification animale doivent pouvoir accéder aux données sur les élevages en toute sécurité grâce à un identifiant unique et l'autorisation explicite des éleveurs pour la récupération de leurs données de mouvements des animaux. Les exploitants doivent également permettre l'accessibilité de ces données pour le suivi des animaux durant toute la chaîne de production.

Les organismes de sélection génétique ont besoin de partager les données de races et d'insémination sur les critères et index génétiques qu'ils suivent. Ils doivent pouvoir suivre le respect des critères de race et analyser les écarts.

Les vétérinaires ont eux besoin de mieux connaître les élevages auprès desquels ils interviennent pour réaliser un meilleur suivi. Ils doivent également pouvoir partager les données de suivi d'intervention, de constats de gestation, de carnet sanitaire, de vaccinations... Les données peuvent également servir à développer des services de téléconsultation et de suivi à distance.

AMÉLIORER LE PARTAGE DE DONNÉES MACHINISTES EN EUROPE

Les données générées par les tracteurs, moissonneuses batteuses, robots de traites, etc. sont enregistrées dans des logiciels embarqués et généralement transmises vers les *clouds* des constructeurs et des concessionnaires. Les usages qui sont faits de ces données sont contractualisés... mais les documents sont difficilement compréhensibles et peuvent amener à une exploitation indue des *data*, sans possibilité pour les exploitants agricoles d'y accéder, encore moins de les partager avec leurs partenaires.

Les *start-ups* veulent accéder à ces données car elles leurs sont utiles pour développer et fournir leurs services aux agriculteurs. Elles doivent pour cela obtenir les autorisations d'accès aux données des agriculteurs, comme le veut la réglementation européenne (Data Governance Act).

Le ministère de l'Agriculture et de la Souveraineté alimentaire doit, lui, accéder aux données des pratiques phytosanitaires des agriculteurs afin d'expérimenter la mise en place du futur registre phytosanitaire. L'expérimentation concerne aujourd'hui les systèmes de gestion parcellaire (FMIS) et pourrait s'étendre à la remontée automatique des données des pulvérisateurs.

Les machinistes et concessionnaires doivent se mettre en conformité avec le nouveau cadre réglementaire (DGA, Data Act) et permettre aux agriculteurs de récupérer l'accès à leurs *data*, pour qu'ils les partagent aux partenaires de leur choix grâce au gestionnaire de consentement/autorisations.

Agdatahub vient justement de signer un accord avec ILVO (Flandres) et DataSpace Europe (Finlande) : pour la première fois, des plateformes de partage de données actives dans le secteur agricole et agroalimentaire s'interconnectent à la façon d'un espace de données transnational afin de faciliter les flux de *data* entre États membres, en réponse à une problématique de cultivateurs de pommes de terre. Grâce à la connexion entre Tritom (DataSpace Europe) et DjustConnect (ILVO), un agriculteur finlandais qui utilise une machine belge AVR peut désormais consulter facilement les données remontant de cet équipement dans son système de gestion agricole. La circulation des *data* se fait simplement entre les plateformes belges et finlandaises. Même chose en France, où des agriculteurs peuvent désormais accéder aux données provenant de leurs machines AVR et les voir remonter dans le logiciel de MyEasyFarm grâce à la mise en réseau entre ILVO et Agdatahub.

CONCLUSION

Par la complémentarité de ses solutions et par sa conformité au cadre réglementaire, l'infrastructure de partage de données Agdatahub traite ces cas d'usage et d'autres, à l'instar des sujets liés au carbone et aux biocarburants, répondant ainsi aux objectifs du pilier « Mieux Nourrir » arbitrés dans la feuille de route « Numérique et données pour la planification écologique » de France Nation Verte. Il s'agit également de répondre à la problématique de la simplification administrative (réclamée, début 2024, par les exploitants agricoles français) et, ainsi, de contribuer à l'amélioration des performances

économique, environnementale et sociétale des exploitations agricoles, comme l'a acté le communiqué de presse¹¹ du ministère de l'Agriculture et de la Souveraineté alimentaire.

Les apports de notre infrastructure de partage de données vont continuer à s'élargir puisque la société Agdatahub basculera d'ici la fin de l'année 2024 sous détention majoritaire publique et deviendra une sorte de bras armé pour déployer les cas d'usage prioritaires identifiés avec les filières, notamment au service de la génétique animale, la filière bovine, les agrocarburants, l'agro-écologie, l'affichage environnemental, le secteur du vin...

¹¹ Communiqué de presse du 16 juillet 2024 du ministère de l'Agriculture et de la Souveraineté alimentaire, <https://agriculture.gouv.fr/le-conseil-dadministration-dagdatahub-vote-le-passage-de-la-societe-en-gouvernance-publique>

Les enjeux du numérique dans la distribution d'électricité

Par Claire WAAST-RICHARD,
Claudine RABILLARD et Yves BARLIER
Enedis

Les réseaux de distribution d'électricité se trouvent, partout dans le monde, au carrefour de nombreuses transitions. La transition énergétique est déjà une réalité et a un impact fort sur ces réseaux qui s'adaptent à leur tour. Le numérique est un des facteurs clés de réussite.

Le déploiement des compteurs communicants, de l'IoT, l'IA, de la simulation... permettent à Enedis de dégager de la performance, d'accompagner ses clients dans leur transition énergétique, et permettent au système électrique et aux acteurs de marché un fonctionnement plus précis. L'augmentation à venir des installations d'énergies renouvelables (EnR) décentralisées, la généralisation de la mobilité électrique, le développement de l'hydrogène impacteront plus profondément la manière de gérer le réseau de distribution.

Le système futur sera beaucoup plus complexe et plus décentralisé. La transition se fera grâce au numérique et aussi grâce à tous les acteurs de la filière, avec une *"smart regulation"* et la mobilisation de tous nos salariés.

INTRODUCTION

Depuis sa création en 2008, Enedis fait partie des quatre principaux distributeurs européens d'électricité, avec 38,8 millions de clients et la responsabilité de la gestion d'un parc d'actifs important : 2 300 postes sources, 1,4 million de km de lignes électriques soit 35 fois le tour de la Terre, 800 000 postes de distribution publique. Enedis, grâce à ses 40 000 salariés, pilote, développe et modernise le réseau électrique de distribution pour garantir à tous ses clients une alimentation électrique de qualité. Enedis joue aussi un rôle clé aux côtés des collectivités et de tous les acteurs locaux. Elle aide à maîtriser les consommations et contribue au développement des mobilités propres.

Le monde change et l'électricité s'impose comme une énergie décarbonée clé. Dans les 12 derniers mois + 12 % en MW d'EnR ont été raccordés sur le réseau d'Enedis, qui accueille 90 % des nouvelles énergies renouvelables, et le nombre de points de charge pour la mobilité électrique s'est accru de 40 %¹. Ainsi la production d'électricité historiquement essentiellement centralisée devient partiellement décentralisée avec des moyens de production intermittents et peu pilotables. Ces transformations dans le système électrique ont un impact fort sur le réseau de distribution qui s'adapte à son tour.

La transformation numérique du réseau est un des facteurs clés de réussite. Si aujourd'hui le système électrique fonctionne déjà grâce au numérique, il reste porteur de perspectives d'améliorations majeures tant pour la performance du réseau de distribution lui-même que pour rendre possible encore plus pleinement la transition énergétique.

¹ <https://data.enedis.fr/pages/accueil/>



Figure 1 : L'organisation du secteur électrique en France (Source : Enedis).

LE NUMÉRIQUE DANS LA VIE DU RÉSEAU ÉLECTRIQUE

La transition énergétique est déjà une réalité sur le réseau public de distribution d'électricité géré par Enedis. Plus de 40 GW de production d'EnR y sont raccordés. En zone rurale cette production dépasse souvent les besoins locaux de consommation. Ainsi, plus d'un tiers de la production EnR est envoyé sur le réseau de transport afin d'être acheminé vers les villes pour y être ensuite distribué par Enedis. Plus de 14 GW d'infrastructures de recharge de véhicule électrique y sont raccordés, répartis sur deux millions de bornes.

Le numérique est omniprésent dans la vie du réseau électrique, de sa planification à son renouvellement, en passant par sa construction, son exploitation et sa maintenance. Les



Figure 2 : Environ 90 % des installations éoliennes et photovoltaïques sont raccordées au réseau de distribution (Source : Enedis).

productions d'EnR et les nouveaux usages de consommation sont pris en compte dès la planification des réseaux. Connaître leur comportement est gage de qualité de la planification. Si celui d'une production éolienne ou photovoltaïque est aujourd'hui bien connu, ce n'est pas encore le cas des recharges de véhicule électrique, dont le profil de consommation s'avère très différent selon l'usage du véhicule. Une autre incertitude pèse sur le déploiement des pompes à chaleur électriques en substitution du chauffage au fioul ou au gaz. L'efficacité énergétique des usages va par ailleurs continuer de progresser. Autant de paramètres qu'il faut évaluer en avenir incertain pour dimensionner le réseau. Cela passe par de la collecte de données, leur analyse, des mises en œuvre d'approches stochastiques ou des simulations des comportements de consommation, de production, du réseau.

Pour l'exploitation du réseau, afin d'anticiper d'éventuelles contraintes (un déséquilibre trop prononcé entre quantité d'électricité injectée et quantité d'électricité consommée conduisant vers un *blackout*), des prévisions locales de production et de consommation sont calculées très régulièrement à l'aval des 2 300 postes sources que gère Enedis. Ces prévisions s'appuient sur des approches d'intelligence artificielle et sont un allié majeur de l'opérateur pour réaliser la gestion prévisionnelle du réseau (du moyen au court terme). En complément des fonctions avancées de conduite (automates) permettent de reconfigurer automatiquement la topologie du réseau moyenne tension en tant que de besoin.

Les compteurs communicants, IoT et traitements de données, sont des atouts supplémentaires pour la performance du réseau. Déjà opérationnels depuis plusieurs années pour les clients de forte puissance, les compteurs communicants ont été généralisés fin 2021 aux 37 millions de clients de puissance plus basse grâce aux compteurs Linky. Linky n'est pas seulement un compteur. C'est aussi un capteur de tension sur le réseau basse tension. L'analyse de ces données permet de détecter des pannes et de les localiser. La recherche de la panne est ainsi plus rapide et la coupure pour le client plus courte. L'analyse des données de mesure (c'est-à-dire issues des compteurs communicants) *a posteriori* permet aussi de réaliser des bilans de qualité de la fourniture pour les clients. Ces données sont aussi utilisées pour détecter des fraudes, identifier des changements de phase nécessaire ou des changements de rapport de transformation dans les postes moyenne tension /

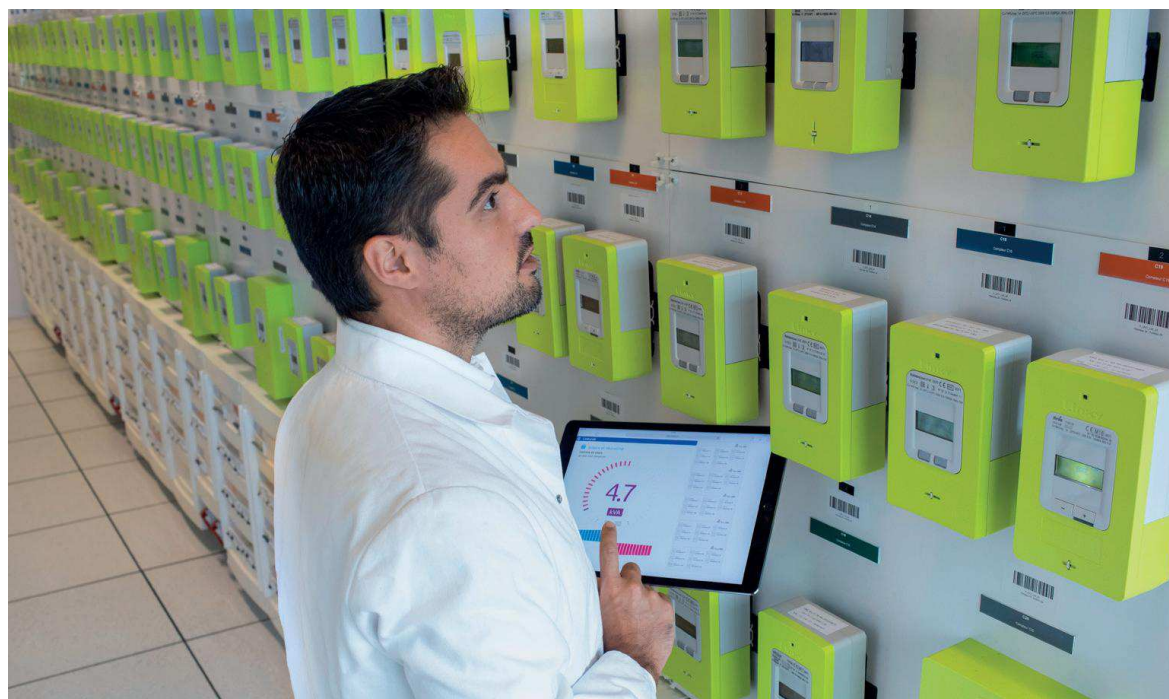


Figure 3 : Linky, une brique essentielle des *smart grids* et de la transition énergétique des parties prenantes (Source : Enedis).

basse tension. En outre, en utilisant l'intelligence artificielle, il est possible d'identifier, quelques jours à l'avance, les risques de panne. Les techniciens d'Enedis peuvent alors intervenir avant la panne.

En complément des compteurs communicants, plusieurs objets connectés sont opérationnels sur le réseau moyenne tension, pour, par exemple, monitorer des transformateurs HTB / HTA, détecter certains défauts des pylônes, rendre les parafoudres, les tourets de câbles ou les groupes électrogènes communicants, anticiper les crues, géolocaliser les équipements. Les analyses des données transmises par ces IoT permettent des actions d'exploitation en temps réel, et d'optimisation de la maintenance ou du renouvellement en temps différé.

Enedis innove également pour les salariés. Des outils de réalité augmentée et de réalité virtuelle ont été déployés pour la formation et l'appui en situation opérationnelle sur le terrain. Pour les activités de bureau, des tâches ont été automatisées afin d'éviter les gestes répétitifs et améliorer la qualité de vie au travail.



Figure 4 : L'IA pour le diagnostic des ouvrages (détection d'une anomalie sur un support de ligne HTA par traitement d'image) (Source : Enedis).

LE NUMÉRIQUE UN ATOUT POUR LA TRANSITION ÉNERGÉTIQUE ET LE FONCTIONNEMENT DU SYSTÈME ÉLECTRIQUE

Les compteurs communicants sont un atout pour la transition énergétique. Outre, les simplifications apportées par les compteurs communicants (relève du compteur et changements de puissance réalisés à distance, facturation au réel, alertes possibles si surconsommation...), ils ouvrent un vaste champ de nouveaux services. Par exemple, Enedis met à la disposition de ses clients, avec leur consentement, leurs données de consommation et de production. Pour les collectivités locales, ces données peuvent être la clé de voûte pour répondre aux enjeux de sobriété, de rénovation, de lutte contre la précarité énergétique, de production locale d'EnR... Le service PrioRéno, développé en partenariat avec la Banque des Territoires et GRDF, illustre ce propos. Il croise des données de consommation et bâtimentaires de sources variées pour produire pour les collectivités et les bailleurs sociaux une cartographie de bâtiments plus ou moins énergivores. Ceci permet de prioriser les chantiers de rénovation à mettre en œuvre. Autre exemple, pour des opérations

d'autoconsommation collective, ce sont les courbes de charge remontées par Linky, avec le consentement des clients, qui permettent de calculer les kWh autoconsommés et ceux revendus.

Les compteurs communicants, sont aussi un atout pour le fonctionnement du système électrique et les acteurs de marché. Avec l'arrivée des EnR et la modification des usages électriques, le système électrique va de plus en plus nécessiter de flexibilités, c'est-à-dire des situations (contractualisées) où des sites de consommation, de production ou de stockage font varier leur puissance à la hausse ou à la baisse en réponse à un signal. Ces flexibilités contribuent à l'équilibre du système électrique. Les EnR, ainsi que d'autres usages flexibles étant majoritairement raccordés au réseau de distribution, c'est Enedis qui fait le « contrôle du réalisé » c'est-à-dire qui s'assure, grâce aux données de comptage et des méthodes algorithmiques brevetés, qu'un service de flexibilité acheté auprès d'un acteur de marché a effectivement eu lieu. D'autres exemples pourraient être cités dont la RECOFlux². Sous cet acronyme, se cache tout un dispositif financier qui sanctionne ou rétribue les acteurs responsables d'un déséquilibre du système électrique (l'équilibre est mesuré par la tension). Les données de comptage et le savoir-faire algorithmique d'Enedis permettent d'identifier quel acteur a aggravé la tension sur le système électrique et quel acteur l'a soulagée.

Premier distributeur d'électricité européen à publier en *open data* des données énergétiques, Enedis place la mise à disposition de données au cœur de sa démarche de transition écologique au service de la société et des territoires. À titre d'exemple, depuis 2015, Enedis publie en libre accès ses données agrégées sur l'énergie électrique, pour éclairer les enjeux de la nécessaire adaptation des usages à la transition énergétique. Ces données ouvertes s'adressent à tous : élus, agents et partenaires des collectivités, producteurs et fournisseurs d'électricité, acteurs de la mobilité électrique et du marché de l'électricité, responsables associatifs et institutionnels, chercheurs universitaires, journalistes, citoyens...

Elles constituent des indicateurs utiles dans des domaines d'action précis et à différentes échelles géographiques : des communes aux frontières nationales en passant par les EPCI, les départements et les régions voire même les adresses pour les bâtiments collectifs. Cartographie du réseau d'électricité, statistiques sur la production et la consommation d'électricité en France et dans les territoires, courbes de charge de consommation et de production à la maille régionale, bilans énergétiques des collectivités et développement local des énergies renouvelables... toutes ces données permettent aux parties prenantes d'analyser l'évolution des usages de l'électricité, de s'impliquer dans les enjeux de la planification écologique et de mieux comprendre la transition énergétique dans sa globalité.

LA TRANSFORMATION VA S'ACCÉLÉRER AU COURS DES ANNÉES QUI VIENNENT

L'augmentation du nombre et de la puissance cumulée des installations EnR décentralisées, la généralisation de la mobilité électrique, le développement de l'hydrogène impacteront de plus en plus profondément la manière de gérer les systèmes de distribution. Le système futur sera beaucoup plus complexe. La modélisation détaillée, en planification comme en gestion prévisionnelle, des installations de stockage d'électricité décentralisées est un défi technique considérable. Dans le même temps, la gestion du système électrique national va évoluer et RTE pour équilibrer le système devra compter sur les moyens de production et consommation qui sont connectés au réseau de distribution. Une coordination forte entre systèmes de transport et de distribution sera indispensable.

² RECOstitution des FLUX.



Figure 5 : Les enjeux du numérique dans la distribution d'électricité (Source : Enedis).



Le système électrique du futur sera beaucoup plus décentralisé. Pour le faire fonctionner, il faudra développer une capacité de *monitoring* et de pilotage à un niveau plus local qu'aujourd'hui. On verra apparaître des postes HTA/BT "*smart*", dotés de capacité de collecter, stocker et traiter des données et de communiquer avec les postes sources et les agences de conduite. Cette intelligence répartie imposera des nouvelles approches de supervision et de gestion à distance. La cybersécurité de nos infrastructures alors qu'elles seront de plus en plus réparties sur le territoire et plus numérique est un défi majeur.

Figure 6 : L'innovation résultera d'un travail collaboratif entre *data scientists* et experts métiers (exemple : solutions de *monitoring* des transformateurs Haute Tension / Moyenne Tension) (Source : Enedis).

L'IA est un levier de performance important. L'usage de l'IA est déjà une réalité à Enedis, par exemple pour prévoir le vieillissement des matériels électriques, optimiser leur maintenance ou leur renouvellement. Elle est utilisée pour évaluer le risque de claquage des câbles souterrains, pour identifier des défauts sur lignes aérienne. Elle est également efficace pour prévoir localement l'ampleur d'incidents à venir quand Météo France annonce des aléas climatiques. On peut alors évaluer les ressources humaines et matériels nécessaires avant même le début d'une tempête. Toutefois des progrès importants sont attendus en auscultation et diagnostic des ouvrages. Ils passeront par un travail collaboratif entre *data scientists* et experts métiers. L'IA générative est aussi dans une dynamique d'innovation foisonnante. Nous travaillons à poser les fondamentaux : l'IA, ce sont d'abord des données accessibles et en juste qualité ; ce sont des compétences ; une gouvernance pour prioriser et garantir conformité, transparence et frugalité des IA, et enfin c'est un mécanisme de maintien en conditions d'intelligence des IA. L'IA est un système apprenant, s'il se trompe il faut le lui « dire ». Le rôle des salariés qui l'utilisent est clé.

Les attentes des salariés évoluent et le digital offre de nouvelles perspectives. Nous cherchons à concevoir des exosquelettes adaptés à nos activités, à mettre à profit les avancées des neurosciences pour prévenir les risques et réduire les accidents, à repenser les solutions d'accès au savoir en situation de travail en tirant parti de l'IA... Nous continuons à développer l'automatisation d'activités répétitives ou pénibles. Nous voulons enfin utiliser la sociologie pour adapter les modes de management aux évolutions des attentes des salariés et renforcer la confiance.

Enfin, pour compléter une liste déjà longue, il faut évoquer les perspectives offertes par des technologies comme la 5G, la *blockchain* ou l'informatique quantique. Chacune de celles-ci permettra d'améliorer la gestion des systèmes de distribution.

CONCLUSION

La transition énergétique est déjà une réalité sur le réseau public de distribution d'électricité et a un impact fort sur le réseau de distribution qui s'adapte à son tour. La transformation numérique du réseau est un des facteurs clés de réussite. Le déploiement des compteurs communicants et des IoT, les traitements de données et l'application d'approches d'IA ainsi que de simulation permettent à Enedis de dégager de la performance opérationnelle, lui permettent aussi d'accompagner ses clients dans leur transition énergétique, et permettent au système électrique et aux acteurs de marché un fonctionnement plus précis. La transformation va s'accélérer. L'augmentation du nombre et de la puissance des installations EnR décentralisées, la généralisation de la mobilité électrique, le développement de l'hydrogène impacteront de plus en plus profondément la manière de gérer le réseau de distribution. Le système futur sera beaucoup plus complexe et beaucoup plus décentralisé. Ce ne se fera pas en un jour et la gestion du réseau a déjà beaucoup évolué au cours des dix dernières années. La réussite se fera grâce au numérique mais aussi dans le cadre d'un écosystème performant, réunissant l'ensemble des acteurs de la filière, avec une "*smart regulation*" et la mobilisation de tous nos salariés.

Les systèmes de transports intelligents coopératifs au service des politiques de mobilité urbaine

Par **Éric MONCEYRON**
Bordeaux Métropole

La connectivité permet désormais d'intégrer le véhicule à son environnement immédiat, et d'apporter une nouvelle aide sécuritaire au conducteur, y compris en faveur d'une cohabitation apaisée avec les autres modes, ou encore de nouvelles opportunités au gestionnaire de voirie et de la circulation. Cette aide est notamment implémentée par les Systèmes de Transports Intelligents Coopératifs (STI-C), basés sur l'échange instantané de données entre les véhicules (V2V) et les infrastructures (V2I).

Après un travail intensif de normalisation, la Commission européenne a impulsé une décennie de projets pilotes successifs pour le déploiement des STI-C. Ils ont permis de se confronter aux réalités opérationnelles et d'atteindre aujourd'hui un stade de mise en production, comme le séminaire final du projet InDiD (Infrastructure Digitale de Demain - juin 2024) a pu en témoigner.

Cet article souligne l'apport des STI-C sur les nœuds urbains du réseau transeuropéen de transport, pour intégrer les véhicules dans une gestion multimodale réactive d'une mobilité globale, notamment dans la périphérie des cœurs d'agglomération où la part modale de la voiture demeure prépondérante.

Le numérique occupe une place importante dans la transition de nos modes de vie, tout particulièrement pour les usages quotidiens de mobilité. Les systèmes de transports intelligents (STI), nés de l'application des technologies de l'information et de la communication aux transports, sont déployés de manière universelle (Wetterwald, 2019). Le véhicule automobile en bénéficie avec la connectivité à bord, poussée par les besoins de marché et la réglementation.

DES VÉHICULES ET INFRASTRUCTURES ROUTIÈRES CONNECTÉS, AU SERVICE DES USAGERS ET GESTIONNAIRES

Une branche s'est particulièrement développée avec les systèmes de transport intelligents coopératifs (STI-C), basés sur l'échange de données entre les véhicules (V2V), les infrastructures (V2I), d'autres parties prenantes (V2X). La normalisation des STI-C a été impulsée par des organismes tels que l'ISO (Organisation internationale de normalisation) et le CEN (Comité européen de normalisation) et l'ETSI (L'Institut européen des normes de télécommunications).

Cette évolution s'est accélérée sous le double effet de l'avènement de la communication sans fil (*wifi* véhiculaire, positionnement satellitaire...) et celui des applications de navigation gratuites sur *smartphone*. Ce dernier s'intègre à bord « sans couture » avec Car Play et Android Auto.

Les gestionnaires routiers se sont intéressés, conjointement avec l'ingénierie automobile (PIARC, 2012), aux potentiels offerts par cette connectivité pour partager les informations entre infrastructure et véhicules. Elle permet d'intégrer le véhicule à son environnement immédiat, y compris dans des zones masquées, et d'apporter une nouvelle aide au conducteur et de nouvelles opportunités au gestionnaire de voirie.

Les normes ont également défini une architecture pour que les services coopératifs soient indépendants des protocoles de communication, et implémentée à bord d'un véhicule, dans une unité de bord de route, ou encore dans un *smartphone* (Bonnin, 2016), ce qui permet d'intégrer un parc de véhicules plus anciens.

UNE DÉCENNIE DE PROJETS PILOTES POUR DÉPLOYER LES STI-C ET ACCOMPAGNER L'HARMONISATION DES STANDARDS EUROPÉENS

En France, le projet pilote SCooP [2014-2017] a permis de tester, sous l'égide du ministère de la Transition écologique, une première vague de déploiement de STI-C, ciblée sur l'amélioration de la sécurité routière, et celle des agents d'exploitation, sur routes nationales et autoroutes (alerte chantiers, opérations d'exploitation en cours, signalisation embarquée des événements dangereux). L'architecture était basée sur des communications wifi ITS-G5, avec l'objectif de déployer 3 000 véhicules sur 2 000 km, puis sur des véhicules de série. SCooP est d'une certaine manière le pivot, après des projets précurseurs, qui a ouvert la voie d'une transformation numérique de la gestion du trafic et de l'information routière.

L'impulsion d'une politique européenne ambitieuse

La Commission européenne a adopté dès 2016 une stratégie en faveur des STI-C (CE, 2016). Elle a ainsi défini le cadre pour un premier déploiement à l'horizon 2019, puis fixé différentes phases jusqu'en 2045, les exigences en matière de sécurité et d'interopérabilité, de protection des données personnelles.

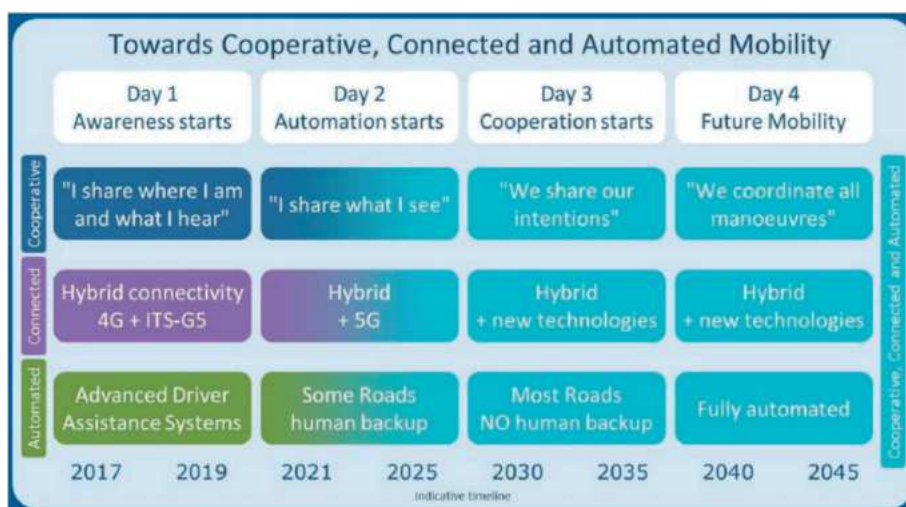


Figure 1 : Stratégie européenne en faveur des STI-C (Source : CE/DG Move).

C2C-CC Roadmap 2024

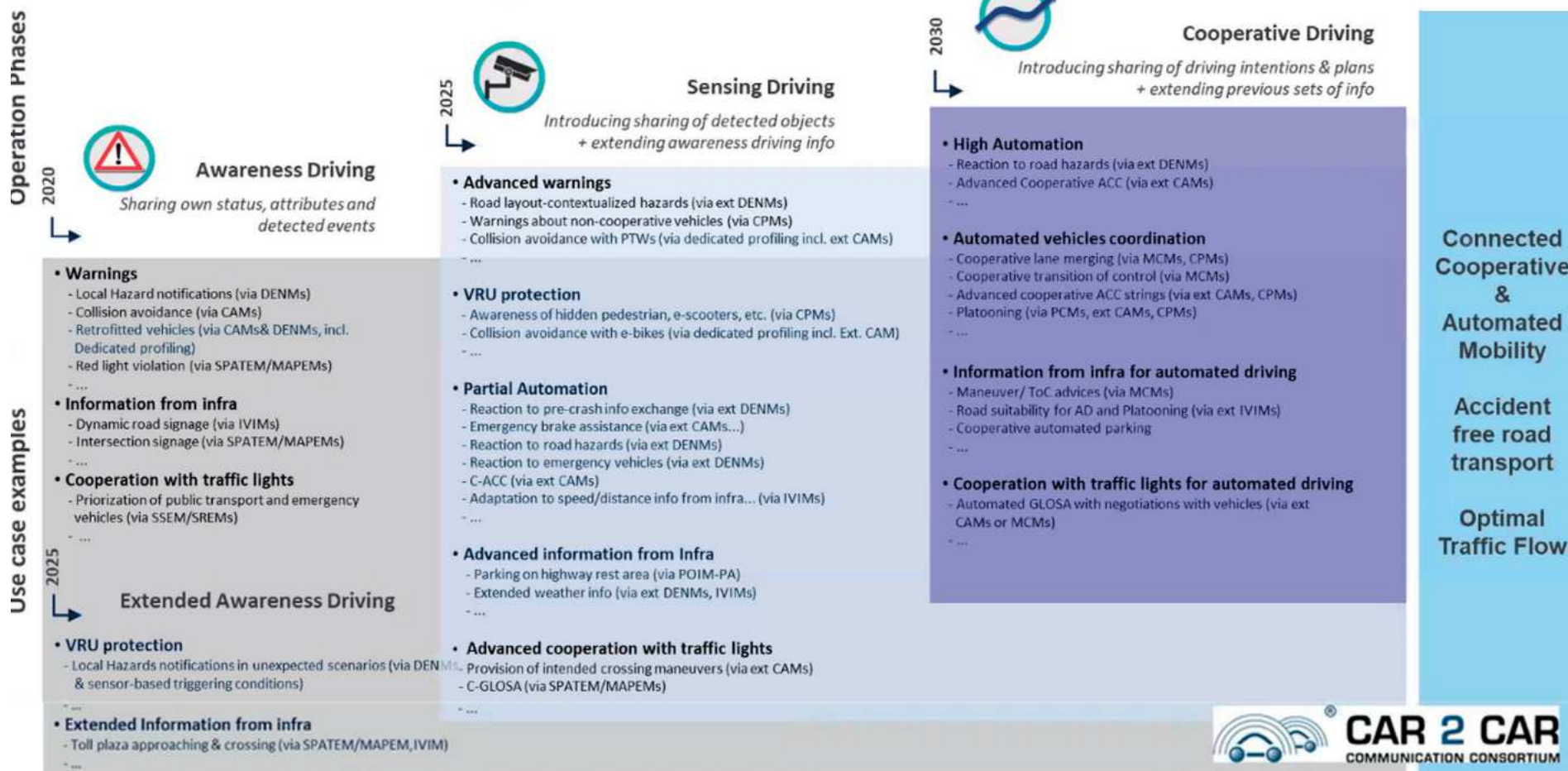


Figure 2 : Feuille de route C2C-CC 2024 (Source : International C-ITS Webinar: C-Roads June 2024).

La Commission a également cofinancé un ensemble de projets (Mécanisme pour l'Interconnexion en Europe), et lancé la plateforme C-Roads, initiative conjointe avec 18 États membres, avec l'objectif d'harmoniser les activités de spécification et de déploiement des STI-C, et favoriser l'interopérabilité au-delà des frontières.

Aujourd'hui, cette plateforme affiche 20 000 km desservis par des services STI-C en courte portée pour environ un million de véhicules équipés, 100 000 km en longue portée ; elle regroupe 50 villes européennes et 11 pays associés. Cette plateforme a aussi initié une coopération étroite avec l'industrie automobile (consortium de communication Car2Car) et l'articulation des feuilles de route respectives.

Le ministère français de la Transition écologique, chef de file des déploiements en France et de l'harmonisation européenne

Le ministère s'est positionné en chef de file de cette plateforme (Esposito, 2024), (Fiorina, 2021), et a coordonné plusieurs projets pilotes successifs (C-Roads France, InterCor, InDiD), qui ont favorisé l'enrichissement incrémental de l'environnement STI-C, réuni différents types d'acteurs (route, ferroviaire, portuaire, forces de l'ordre et de secours...) et abouti à un socle technologique de référence¹, avec :

- Une architecture réseau hybride (ITS-G5 et cellulaire), articulée autour du nœud national, permettant d'adresser un grand nombre de cas d'usage aux utilisateurs finaux, de manière agnostique.
- Un dispositif de signature des messages (Public Key Infrastructure), permettant de constituer un domaine de confiance pour garantir l'authentification des messages échangés, et venant ainsi renforcer l'anonymisation des véhicules par utilisation de pseudonymes qui changent régulièrement.
- L'application *smartphone* « Coopits », qui permet également un affichage superposé des informations sur le navigateur numérique de l'utilisateur (Coyote, TomTom, Mappy, Waze, Google Maps...).
- Un catalogue de services (13 familles, 101 cas d'usage).

Cette évolution a été orchestrée par le comité technique national, associant de nombreux partenaires et groupes de travail spécialisés.

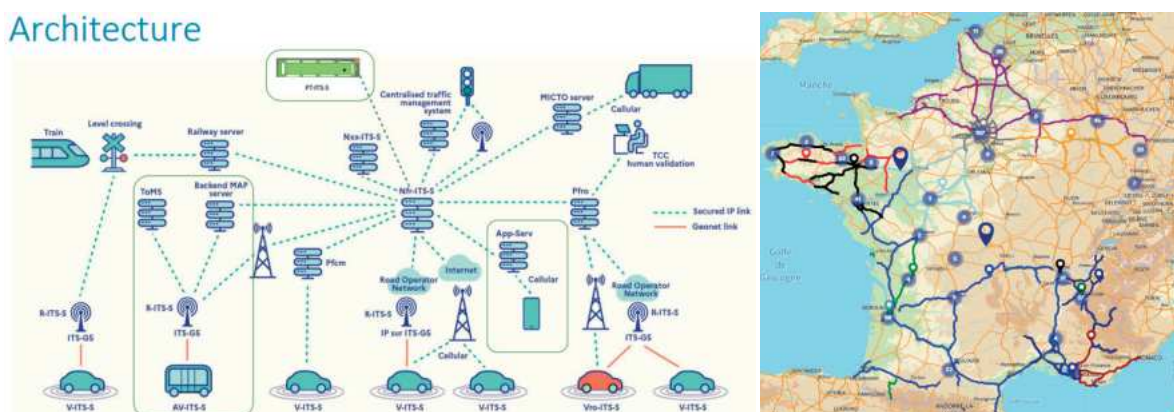


Figure 3 : Architecture et périmètre de déploiement Coopits
(Source : projet InDiD, 18^e baromètre AXA, 2022).

¹ <https://c-its.developpement-durable.gouv.fr/>

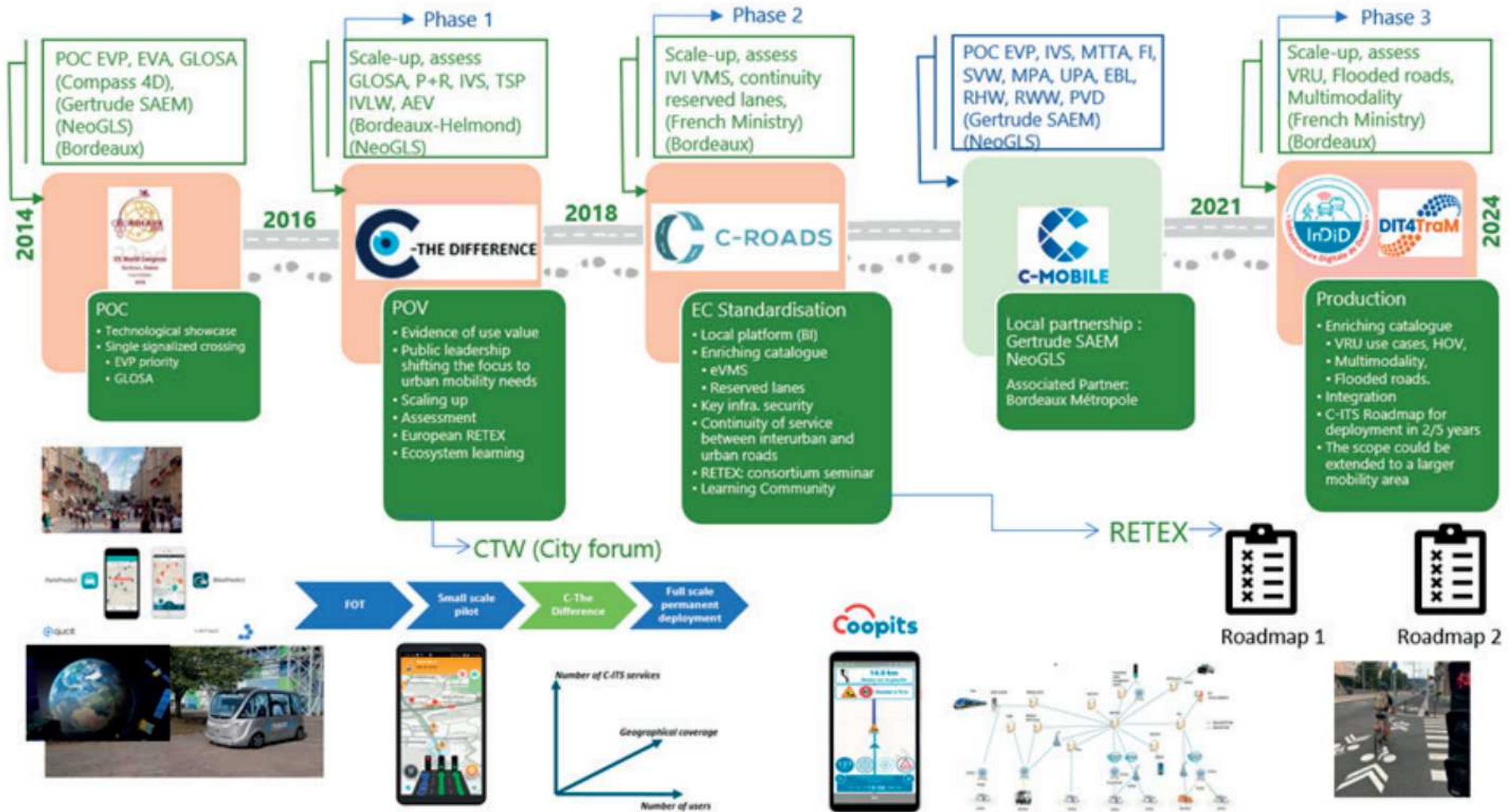


Figure 4 : Expansion du site pilote bordelais et augmentation progressive du TRL - Technological Readiness Level (niveau de maturité technologique) (Source : Bordeaux Métropole).

Le ministère affiche aujourd’hui 5 000 km de réseaux routiers desservis par des services STI-C en courte portée (579 unités bord de route), 15 000 km en longue portée *via* Coopits. Les cas d’usage déployés concernent en majorité les enjeux de sécurité routière interurbaine (alertes diverses, événements routiers, signalisation à bord, limitation de vitesse...).

L’étude socio-économique (Chakir, 2023) met en exergue que seul un scénario de déploiement des STI-C en communication hybride génère une valeur actualisée nette positive et donc bénéfique pour la société (gains évalués dans une fourchette de 500 à 2 000 M€ sur la période 2022-2052). Un tel scénario maximise les effets des services activables (I2V, V2V, V2I2V) et leur périmètre, donc les bénéfices générés (diminution de l’accidentologie, amélioration du trafic...).

Cette étude souligne également que l’utilisation de *smartphones* comme outil d’assistance à la conduite permet la réduction d’externalités négatives. En 2022 près de 45 %² des conducteurs utilisent le GPS de leur *smartphone*. Cet usage devrait encore augmenter, être plus précis, notamment avec l’avènement de la 5G : hypothèse d’une croissance de 45 % (2022) à 60 % (2030) du taux des trajets sur lesquels le *smartphone* est utilisé, puis une stagnation après 2030.

LES STI-C, SOCLE TECHNIQUE D’UNE MOBILITÉ URBAINE INTELLIGENTE

En secteur urbain, “C-the difference” [2016-2018] fut le projet pivot, avec un consortium public/privé européen autour de deux territoires d’expérimentation : Bordeaux Métropole et la ville de Helmond³. Le déploiement à grande échelle a été facilité par le choix de développer une application *smartphone* (“C-the difference App”), jouant le rôle d’une station STI embarquée.

Ce projet a favorisé le passage d’une preuve de concept technologique à une preuve de valeur d’usage, en déplaçant le curseur vers les besoins de mobilité en lien avec la politique publique des déplacements urbains. Les villes ont sélectionné les cas d’usage (*cf.* Figure 5 page suivante) permettant de viser un impact significatif sur l’efficacité du trafic, la réduction⁴ de la consommation de carburant et des émissions polluantes en favorisant une écoconduite pour le franchissement des carrefours, en diminuant les arrêts aux feux, le temps et les kilomètres dans la recherche d’une place de parking.

Ce site pilote STI-C urbain a été enrichi et robustifié au cours des deux projets « C-Roads France » et « InDiD ». La montée en TRL⁶ s’est faite progressivement, jusqu’à un stade de mise en production aujourd’hui (*cf.* Figure 4 page précédente). Ce site est adossé au socle technologique de référence.

Intégration des véhicules dans une mobilité multimodale et raisonnée

L’application Coopits a été déployée sur l’agglomération bordelaise. Elle permet de mieux connecter les réseaux gérés par les différents opérateurs, et assure ainsi une continuité de service entre la voirie urbaine, la rocade et la route interurbaine, à l’échelle d’un nœud

² 18^e baromètre AXA (2022).

³ Aire urbaine de Eindhoven – et pour la partie française : Bordeaux Métropole, NeoGLS, Cerema et l’Université Gustave Eiffel.

⁴ Jusqu’à 10 % avec un taux de pénétration significatif au-delà de 40 % - IRF Task Force (2021), “Preliminary report, ITS for Climate Impact Mitigation”, 17 pages.

<p>A – Collecte de données</p> <p>A1 – Collecte de données trafic</p> <p>A3 – Remontée d'événements déclarés manuellement</p> <p>B – Alerte chantiers</p> <p>B1a – Alerte fermeture partielle ou complète d'une ou plusieurs voies</p> <p>B1b – Alerte fermeture programmée d'une route ou autoroute</p> <p>C – Signalisation embarquée</p> <p>C2 – Information sur la vitesse limite dynamique dans le véhicule</p> <p>C3 – Signalisation embarquée (PMV embarqué)</p> <p>D – Événements inopinés et dangereux</p> <p>E – Information de trafic et reroutage</p> <p>F – Stationnement, parc relais et multimodalité</p> <p>F1a – Information sur la localisation des parkings en ouvrage, leur disponibilité et les services associés - aires de covoiturage</p> <p>F5 – Information P+R pour du transfert modal</p> <p>G – Carrefours à feux</p> <p>G1a – GLOSA (conseil de vitesse optimale pour passer au vert)</p> <p>G1b – TTG (temps au vert)</p> <p>G2 – Priorité au feu requise par un véhicule désigné - Véhicules de secours (POC 2019) - Vélos connectés (POC 2023)</p> <p>G3 – Non-respect de la priorité au feu : alerte au véhicule ego (POC 2018)</p> <p>G8 – Facilitation et sécurisation des déplacements d'usagers spécifiques à un carrefour à feux (spécifications) - Application aux personnes aveugles et malvoyantes (POC 2024)</p> <p>H – Gestion de trafic</p> <p>H4 – Gestion dynamique de voies - Voies réservées - Voie covoiturage - Navette bus sur BAU</p> <p>H9 – Routes inondées</p> <p>I – Usagers vulnérables</p> <p>I6 – Piéton traversant la voie sur passage protégé sans signalisation tricolore alerte aux véhicules à l'approche</p> <p>J – Optimisation de transport de fret et logistique</p> <p>K – Passages à niveaux</p> <p>L – Forces de l'ordre</p> <p>M – Services de paiement</p>

Figure 5 : Liste des services STI-C sur Bordeaux Métropole (2024)

urbain de l'arc atlantique. Elle intègre près de 30 cas d'usage actifs, support aux politiques publiques sur différents secteurs :

- en cœur d'agglomération : renforcement de la sécurité des usagers les plus vulnérables, tels que les cyclistes et les piétons, dans leur cohabitation avec les véhicules motorisés notamment aux intersections et le soutien au développement des modes actifs ; cela vient compléter les cas d'usage de conseil de vitesse en amont des carrefours à feux ou encore de demande de priorité pour des véhicules désignés ;

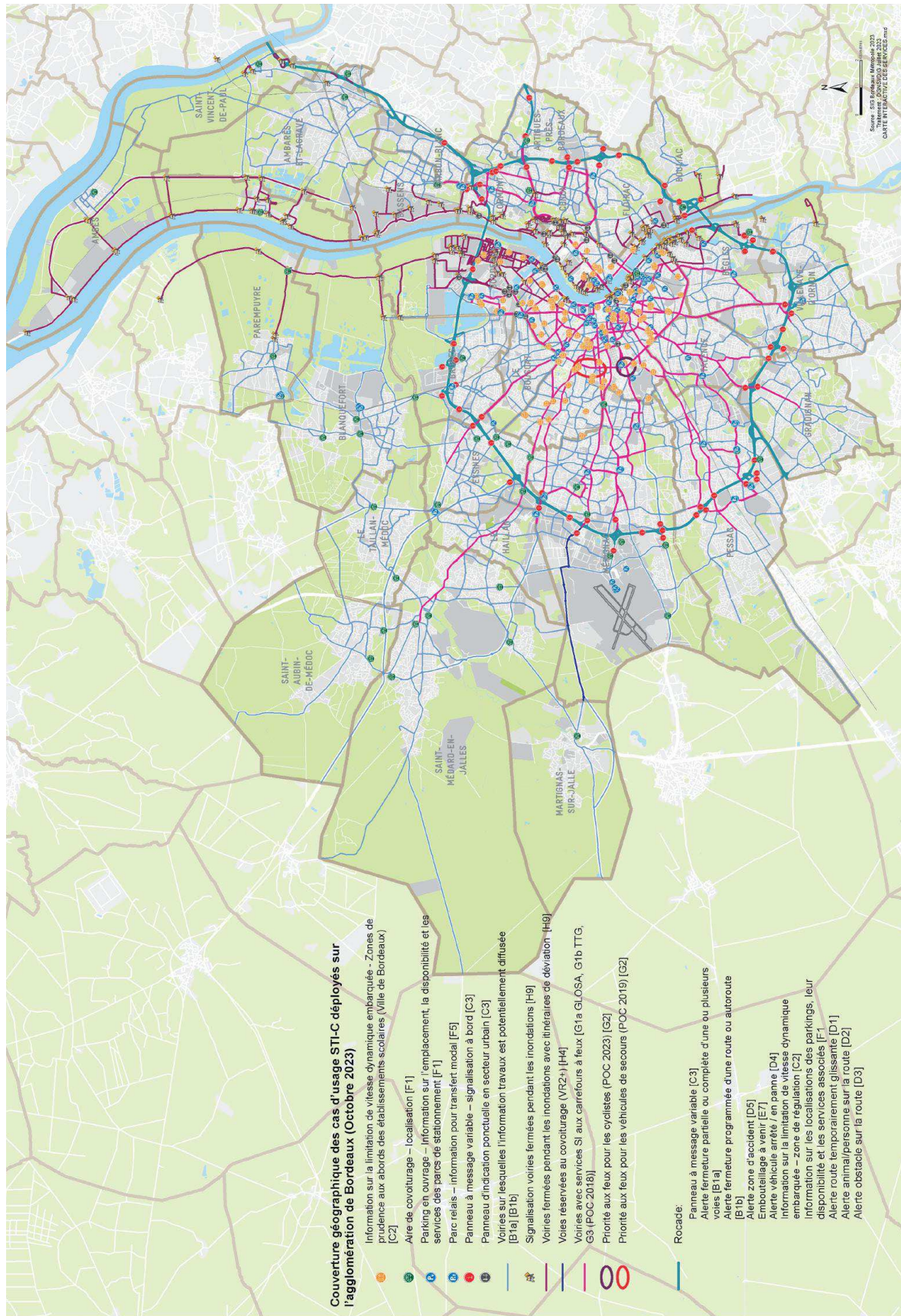


Figure 6 : Couverture géographique des services STI-C sur l'agglomération bordelaise (2024).

- sur la rocade et pénétrantes : amélioration de la sécurité des usagers et des agents, optimisation de l'information aux usagers en temps réel ;
- en couronne périurbaine : incitation au report modal vers le *Mass Transit* via les parkings relais, information sur l'accès aux aires de covoiturage et voies dédiées - sécurité routière.

Ces services intègrent les véhicules dans une gestion multimodale réactive de la mobilité urbaine. Un objectif plus global est aussi d'améliorer la gestion d'événements spécifiques comme l'information sur les tronçons routiers inondés lors des crues estuariennes.

C'est aussi une nouvelle approche de l'utilisation et de la gestion de l'infrastructure urbaine à partir d'occurrences ou d'événements numériques en temps réel (Picon, 2016). Elle vient compléter les systèmes d'aide à la gestion de trafic qui, au-delà de la régulation des flux et notamment sur des secteurs plus périphériques, sont en mesure de délivrer des services numériques à bord des véhicules (alerte travaux ou danger signalé⁵, cycliste à contre sens, gestion dynamique de voies réservées...).



Figure 7 : Cas d'usage F5 d'information dynamique en faveur d'un transfert modal sur P+R (Bordeaux).

À titre d'illustration, le cas d'usage F5 vise à informer les automobilistes approchant des parkings relais sur la disponibilité, les délais avant le prochain tramway, les comparaisons de temps de trajet entre différents modes de transport et les incitations à utiliser les transports publics. Il est déployé dans la zone d'influence du P+R, avec une distinction entre les secteurs éloignés et proches, où les messages encouragent respectivement le transfert modal et fournissent des informations pratiques sur les places disponibles et les fréquences de transport.

Le cas d'usage H9 consiste à alerter les usagers des tronçons de voiries inondés, après déclenchement du dispositif consécutif à une crue estuarienne, et informer sur les itinéraires de déviation. Les citoyens peuvent ainsi s'approprier les risques du territoire de manière plus précise, renforçant ainsi leur résilience et les liens sociaux.

À partir des tests réalisés sur l'agglomération bordelaise (2021 et 2023), des recommandations sont proposées. Par exemple, pour G1a (GLOSA) afin d'améliorer le comportement des conducteurs : distance d'activation en fonction du temps de cycle de feux et du taux

⁵ Eurométropole Strasbourg – SIRAC - Évènement final du projet InDID 12 juin 2024.

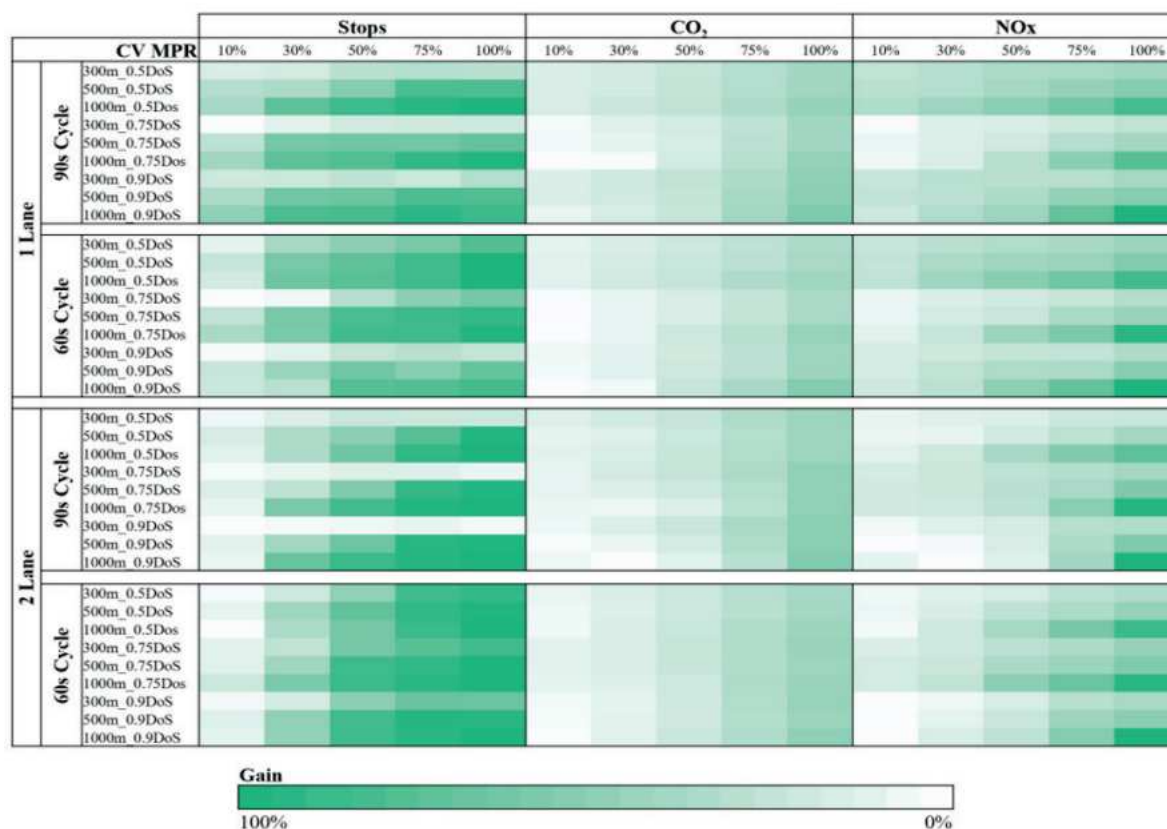


Figure 8 : Impact du GLOSA sur les performances de circulation et environnementales (Laharotte et Bhattacharyya, 2021).

de saturation de la circulation, bénéfique sur l'efficacité environnementale en fonction du taux de pénétration du service.

Des services de plus haut niveau peuvent être envisagés, renforçant l'infrastructure servicielle en support des politiques de mobilité urbaine. Nous pouvons considérer avoir atteint l'étape 2 d'un déploiement coordonné entre C-ROADS et C2C-CC (cf. Figure 9 page suivante). Bordeaux Métropole envisage de poursuivre autour des enjeux de sécurisation des modes doux, de multimodalité et de contribuer à l'accompagnement pédagogique nécessaire à l'instauration d'un périmètre ZFE d'intermodalité.

LES LEVIERS À ACTIONNER

Le groupe de travail national orchestré par le ministère a classifié l'ensemble des cas d'usage selon leurs besoins fonctionnels (latence, précision...), et souligne la nécessité d'une communication hybride avec un volet connectivité de courte portée stabilisé.

Deux technologies restent aujourd'hui concurrentes sur la courte portée et opérant dans la même bande de fréquences (5,9 GHz) : l'ITS-G5 (wifi véhiculaire) et le C-V2X (Cellular-V2X, alternative standardisée dérivée des réseaux cellulaires 4G, et bientôt 5G-V2X). Certains gestionnaires routiers et constructeurs automobiles (Volkswagen) ont fait le choix de la première d'entre elles, disponible et plus mature. D'autres OEMs affichent une préférence de plus en plus marquée pour la seconde technologie, à l'instar de la plateforme automobile en France. Les deux nécessitent le déploiement d'unités bord de route. Le fait que le choix d'une technologie n'ait pas été tranché par l'exécutif européen, maintien de l'incertitude sur les feuilles de routes respectives. La cible d'un taux de pénétration élevé des services STI-C requiert une coordination renforcée des différents acteurs.

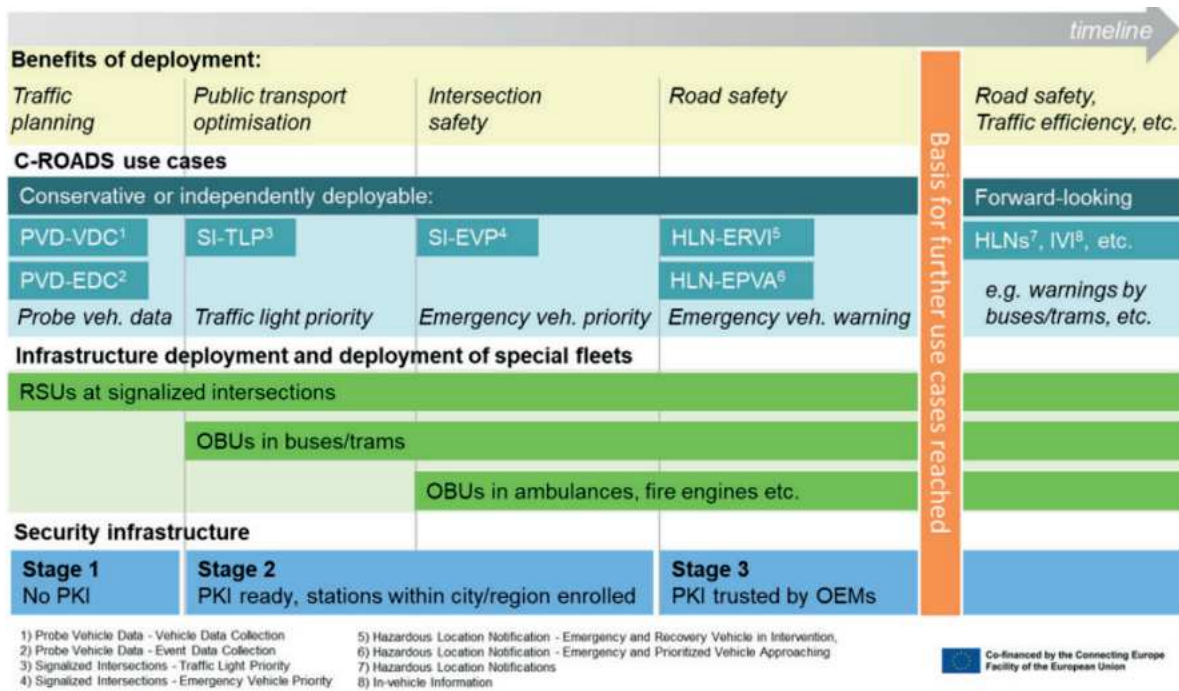


Figure 9 : Synergies de déploiement progressif des STI-C urbains (Ruehrup *et al.*, 2023).

Le programme européen d'évaluation des nouveaux véhicules (EuroNCap) commence à intégrer depuis 2023 l'apport de la connectivité en faveur de la sécurité⁶. Dès 2025 l'accent est mis sur l'utilisation des technologies avancées pour améliorer la sécurité des véhicules mais aussi celle des usagers vulnérables (piétons, vélos). C'est donc un nouveau levier qui devrait inciter les OEMs à accélérer le déploiement.

Un axe complémentaire, à harmoniser, vise la création d'un espace commun pour partager les données collectées à partir des capteurs embarqués, et constituer des événements après qualification/agrégation.

Le déploiement des STI-C augmente également les surfaces d'attaque cyber des systèmes. La directive⁷ NIS 2 portant notamment sur la sécurité informatique et la protection des données, marque un tournant avec des exigences renforcées notamment pour les métropoles. La transposition en droit français est prévue pour octobre, levier réglementaire qui nous oblige vis-à-vis des exigences des politiques de certification et de sécurité européennes des STI-C.

CONCLUSION

Les STI-C ont tous les atouts pour atteindre un taux de pénétration suffisant, via une communication hybride unifiée, interopérable, ouverte et cybersécurisée. Ce potentiel permet d'envisager l'industrialisation et l'élaboration de nouveaux cas d'usage de haut niveau. La révision récente de la directive européenne STI et ses règlements renforcent ces objectifs. La position adoptée en France est de poursuivre la dynamique engagée dans le cadre du nouveau projet SCALE sélectionné par la Commission européenne mi 2024.

⁶ Règlement UE « SRTI » (*Safety related traffic information*) (2013/886) qui porte sur les événements de sécurité routière et sur les échanges de données.

⁷ *Network and Information Security*.

BIBLIOGRAPHIE

- ATEC ITS FRANCE (2015), « Mobilité 3.0, Ensemble pour la mobilité intelligente », Livre vert, septembre, 60 pages, <https://atec-its-france.com/production/mobilite-3-0-ensemble-mobilite-intelligente/>
- BONNIN J.M. (2016), « Standards & ITS : vers un écosystème ouvert », *revue TEC*, n°231, pp. 44-45.
- BONNIN J.M. & COUTURIER C. (2020), « Les infrastructures de communication », *revue TEC*, n°245, pp. 22-23.
- CE (2016), « Stratégie en faveur des systèmes de transport intelligents coopératifs (STI-C), jalon d'une mobilité coopérative, connectée et automatisée », Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen, et au comité des régions, novembre, Bruxelles.
- CHAKIR J. (2023), « Résultats de l'analyse socio-économique de InDiD », C-rapport projet InDiD, réf. 2.3.6.2c, DGITM.
- DG Move (2018), Study Report, Consortium, "C-The Difference Pilot - Beyond Traffic Jams", Bruxelles.
- ESPOSITO M.C (2024), « La France chef de file dans le déploiement des C-ITS », *revue TEC*, n°261, pp. 34-35.
- FIORINA C. (2021), « Les systèmes de transport intelligents coopératifs ou C-ITS », *revue TEC*, n°250, pp. 38-39.
- HAËNTIENS J. (2021), *Smart city, ville intelligente : quels modèles pour demain ?*, Paris, La documentation française.
- LAHAROTTE P.A. & BHATTACHARYYA K. (2021), "Performance analysis and recommendations on use cases C2 and G1", C-Roads project report, ref. 2.3.7.7, UGE.
- MONCEYRON E. (2015), « Bordeaux Métropole : territoire d'innovation et d'expérimentation », *revue RGRA*, n°932.
- PIARC - Technical Committee B.2 Road Network Operations (2012), "The connected vehicle", technical report 2012R02F.
- PICON A. (2016), « Une autre conception des infrastructures de la mobilité se dessine », *revue TEC*, n°229, pp. 4-5.
- RUEHRUP S., CONCEIÇÃO L., MONTENEGRO J. & MECKEL P. (2023), "The Chicken and the Egg – Perspectives of C-ITS Deployment", 1. ASFINAG, Austria, 2. ARMIS, Portugal, 15th ITS European Congress, Lisbon, Portugal, May.
- WETTERWALD M. (2019), « La normalisation des systèmes de transport intelligents », *Annales des Mines – Enjeux numériques*, mars, pp. 75-82.

Plaidoyer pour les grandes oubliées : les infrastructures publiques de partage de données

Par Laura LÉTOURNEAU
France Nation Verte

Et si les infrastructures du futur étaient moins celles dont les technologies sous-jacentes sont spectaculaires que celles qui nous aident à relever les défis... du futur ? Les infrastructures de partage de données organisent la circulation fluide et sécurisée d'informations au sein d'un cercle donné d'acteurs. Transition écologique, gestion des pandémies, amélioration du système éducatif, sécurité... De telles infrastructures sont en réalité nécessaires dès lors que de nombreux acteurs publics et privés doivent travailler ensemble pour résoudre des problèmes de société. Quoi de plus important ? Le sujet est pourtant victime d'une asymétrie d'attention délétère au profit de technologies plus « sexy » comme l'IA. Le présent travail explique en quoi ces infrastructures sont indispensables et fait émerger un guide terrain, collectif et ambitieux pour aider à les porter, avec la conviction que l'on peut aller collectivement plus vite lors des 3 prochaines années que lors des 20 dernières.

DÉMARCHE : FAIRE ÉMERGER DES LIGNES DIRECTRICES TERRAIN, COLLECTIVES ET AMBITIEUSES POUR PASSER UN CAP

Quand en 2016 avec Clément Bertholet, nous avons écrit « Ubérisons l'État ! Avant que d'autres ne s'en chargent » dans la lignée des travaux d'Henri Verdier et de beaucoup d'autres personnes, on nous a dit que c'était déjà dépassé et qu'il aurait fallu écrire « Blockchainisons l'État ». En 2020, pendant les premiers mois du Covid, alors que nous travaillions jour et nuit pour mettre sur pied les nombreuses infrastructures de partage de données nécessaires au dépistage, au *contact tracing*, à la gestion de la logistique des vaccins, à leur traçabilité, à la gestion des potentiels effets secondaires, à l'envoi de masques et de respirateurs, à l'identification des lits d'hôpitaux disponibles et au transfert de patients, nous avons dû répondre en parallèle à de nombreuses questions sur la façon dont on utilisait l'intelligence artificielle pour gérer la pandémie. Le rôle que l'IA pouvait jouer dans les problèmes critiques à résoudre à moyen terme était pourtant minime ou nul.

Les infrastructures de partage de données ne sont pas particulièrement attrayantes. Comme les infrastructures physiques (les réseaux d'eau, d'égout, internet...), leurs technologies sous-jacentes sont rarement spectaculaires. Leurs cas d'usage sont diffus. Leur exécution est complexe et de long terme. De la même façon qu'un élu préfère inaugurer un bâtiment (un hôpital, une école, un incubateur de *start-ups*) plutôt qu'un réseau d'égouts,

les politiques, la presse et les capitaux sont davantage attirés par un service numérique concret et innovant plutôt que par une infrastructure numérique peu tangible.

Les infrastructures de partage de données sont pourtant essentielles pour concevoir et mettre en œuvre les politiques publiques. Il s'agit de gérer le dépistage du Covid (avec SI-DEP) ou d'améliorer la coordination ville-hôpital et la prévention en santé (avec Mon espace santé). D'améliorer le revenu des agriculteurs qui adoptent des pratiques environnementales vertueuses (avec Agdatahub). De gérer le remplacement des professeurs absents ou la gestion des bâtiments scolaires (avec l'infrastructure de l'éducation). Ou encore de mettre en place l'affichage environnemental, de gérer efficacement les mobilités et d'accélérer la rénovation des logements en ciblant efficacement les passoires thermiques, en estimant l'effet d'une rénovation sur le DPE ou en luttant contre la fraude (avec respectivement la base Empreinte et la base de Produits réels, EONA-X et moB, l'espace numérique du logement).

Une infrastructure de partage de données structure, agrège, met en lien des données, et *via* ces dernières, des acteurs publics et/ou privés qui peuvent alors échanger, se comprendre, faire société pour changer les systèmes dysfonctionnels dans l'espoir d'avancer ensemble vers un destin commun. Quoi de plus important ? Il s'agit de retrouver notre humanité : ce qui rend un projet numérique exaltant, ce n'est pas la technologie sous-jacente, ce sont les objectifs au service desquels elle se met.

Les oppositions aux infrastructures de partage de données sont rarement explicites. Le plus souvent, les discussions sont dispersées et ne se déroulent pas au bon niveau. En gros : c'est mou. Lorsqu'elles le sont, elles consistent moins en une remise en cause du besoin qu'en des craintes sur le bon cadrage ou la bonne exécution du projet du type : « on écrase une mouche avec un marteau », « ça ne passera pas la Cnil », « on n'a pas les moyens » ou « on fera bien moins que le privé ».

Or, comme pour la gestion des crises Covid ou climatique, la question à se poser est moins « est-ce que l'on va y arriver ? » que « comment va-t-on y arriver ? » ou en tout cas, faire au mieux. D'autant plus que comme le montrent les exemples réussis (impots.gouv.fr, SI-DEP...), c'est possible : le plus souvent, on n'a juste pas assez ou pas assez bien essayé.

C'est parce que le décalage entre les enjeux colossaux associés aux infrastructures de partage de données et l'amateurisme avec lequel elles sont souvent portées ne nous est plus supportable, parce que l'on pense que l'on doit faire mieux, que l'on peut faire mieux, et que c'est notre responsabilité à nous – les quelques personnes qui travaillons de ces sujets – d'y contribuer, que l'on a tenté de faire émerger un guide pratique terrain, collectif et ambitieux dans l'espoir de passer un cap.

POURQUOI ? : RÉURBANISER LE BIDONVILLE POUR ÊTRE À LA HAUTEUR AUX ENJEUX DU SIÈCLE

Le guide commence par décrire l'état des lieux et la raison d'être du travail réalisé. Sorte de catharsis collective, cette partie est essentielle pour partir sur de bonnes bases.

Avant de zoomer sur les infrastructures de partage de données, elle se penche sur le numérique au sens large : est-il vraiment nécessaire pour répondre aux enjeux du siècle ? Ou bien ce n'est pas le sujet car, d'une part, « ce n'est pas le cœur de métier » (« Notre priorité, c'est de décarboner l'industrie ») et, d'autre part, le numérique ne répond pas à notre place à l'essentiel : que veut-on profondément ? Par exemple, dans l'écologie, à quoi souhaite-on que notre vie ressemble demain ? Ou pire, le numérique est carrément nuisible : fuite de données sensibles, fractures numériques, empreinte environnementale, technosolutionnisme, amélioration effrénée de la production, diminution de la résilience, fausses informations, bulles de filtres... Les maux sont graves. Et à la fois comme l'argent,

le numérique est (malheureusement ?) un moyen indispensable pour mettre en œuvre efficacement les politiques publiques (par exemple pour décarboner l'industrie). Au-delà de permettre, il a même la capacité à déclencher des changements plus profonds : modifier les relations médecins/patients, passer de la propriété à la location...

C'est le moment d'embrasser la complexité : inaugurer le navire, c'est inaugurer le naufrage. Dès lors que le coût/bénéfice est positif, il faut inaugurer le navire tout en minimisant de façon ultra proactive les risques de naufrage, en développant le numérique dans un cadre de valeurs éthique, humaniste et citoyen.

Les infrastructures publiques de partage de données jouent un rôle pivot. Elles sont appelées *digital public infrastructures* au niveau onusien et *data spaces* au niveau européen. Elles organisent le partage fluide et sécurisé de données qui n'ont pas vocation à être ouvertes du fait de leur sensibilité au sein d'un cercle fermé d'acteurs. Elles urbanisent l'échange de données, au lieu de laisser chaque acteur bricoler dans une sorte de bidonville où chacun tire ses propres lignes d'électricité et se débrouille pour gérer ses déchets. Elles sont la couche haute de la « plateforme publique ». À ce titre elles s'appuient sur les règles d'éthique, d'interopérabilité, de sécurité, les bases de données et les outils d'identification et de gestion du consentement construits par le public dans une logique de « commun » (fondations du bâtiment, voir la Figure 1). À l'inverse, les services numériques à valeur ajoutée métier et les observatoires (murs et toit), principalement développés par l'externe, s'appuient sur elles pour faire circuler les données au service de cas d'usage d'intérêt général.

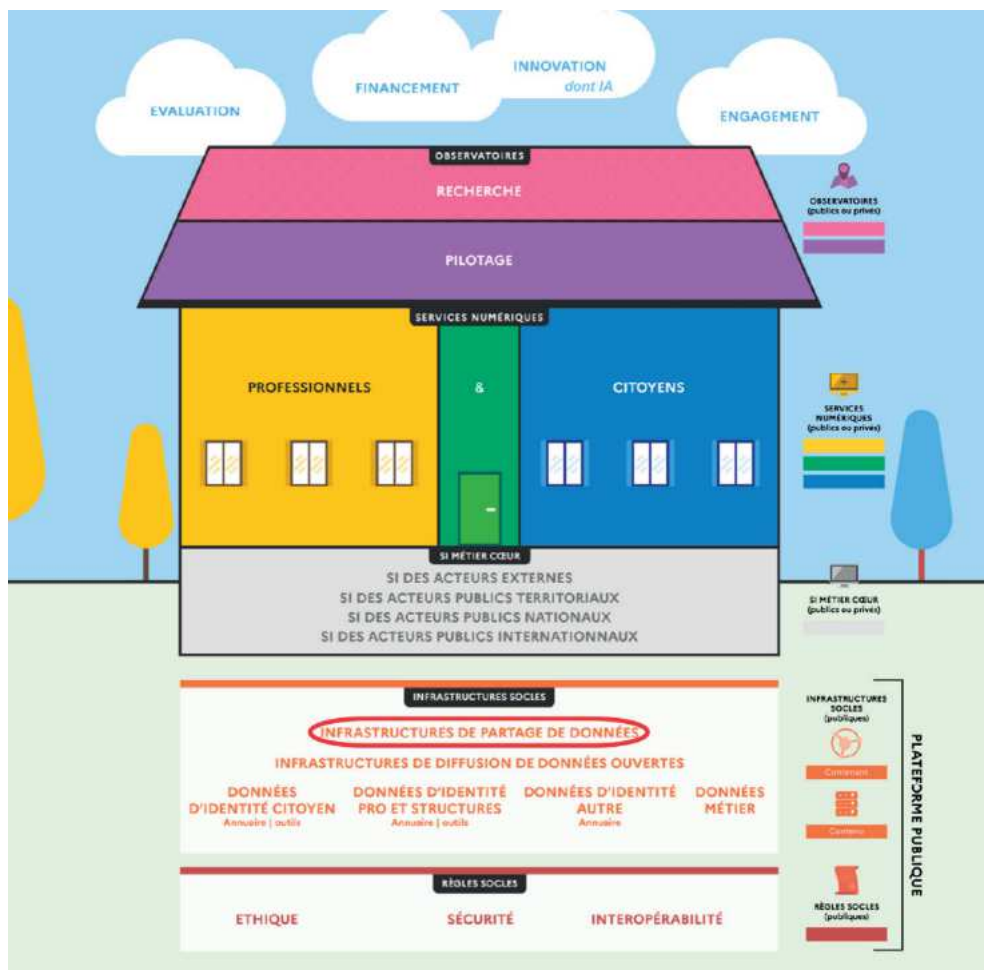


Figure 1 : La doctrine de « plateforme publique » (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

Leurs bénéfices directs et indirects sont colossaux.

Pour l'efficacité métier et budgétaire des pouvoirs publics. De nombreux rapports d'inspection pointent les défaillances en partage de données comme un élément qui nuit considérablement à la mise en œuvre efficace et collective des politiques publiques. Y remédier est la raison d'être première de ces infrastructures. Par ailleurs, même si cela reste malheureusement souvent compliqué à estimer avec précision, elles représentent un investissement dont le coût, bien que supérieur à celui d'un service numérique classique, est souvent de plusieurs ordres de grandeur inférieur aux économies qu'elles permettent de réaliser. Mon Espace Santé contribue, par exemple, à diminuer significativement la redondance d'actes en radiologie et en biologie, dont le surcoût est estimé entre 1 et 5 Mds€ par an.

Pour l'efficacité environnementale et économique des entreprises, les infrastructures sont la colonne vertébrale de la simplification administrative. Elles permettent aux entreprises de communiquer leurs données une bonne fois pour toute aux pouvoirs publics. C'était notamment une demande récurrente des agriculteurs lors de la crise de début 2024. Elles permettent aussi la sobriété environnementale. Les rappels de vaccins et d'examens de prévention dans Mon Espace Santé contribuent par exemple à diminuer le recours au soin, et donc l'empreinte environnementale de la filière. Elles accélèrent enfin la réindustrialisation. Il est par exemple estimé qu'elles permettent aux entreprises d'augmenter leur satisfaction client de 15 % et de réduire leurs coûts de 11 % par an les premières années.

Pour un développement innovant et souverain du numérique. Les infrastructures sont des projets « totem » qui permettent de tirer le déploiement de toute la plateforme publique : lorsqu'un hôpital se connecte à Mon Espace Santé, il est par exemple obligé de renforcer ses règles d'accès, ce qui améliore sa cybersécurité globale. Elles permettent par ailleurs la création de services numériques innovants, qui n'auraient pas vu le jour sinon. Par exemple, la circulation des données collectées dans ProNote permettra aux *start-ups* de l'EdTech de proposer de nouveaux services, notamment à base d'IA. Enfin, comme expliqué, la maîtrise de ces infrastructures est essentielle à notre souveraineté « démocratique » : le public (et non le privé) doit piloter ces briques car elles sont sur le chemin critique de la mise en œuvre des politiques promises. Mais elle est aussi nécessaire à notre souveraineté « technologique » : la maîtrise des données alimentant les modèles d'IA aidera par exemple au développement d'une filière privée française et européenne. En retour, la souveraineté « technologique » renforce la souveraineté « démocratique », le public ayant besoin d'industriels efficaces et souverains en prestataires, pour l'aider à construire en marque blanche les infrastructures.

Le développement de ces infrastructures est donc nécessaire. Il est par ailleurs réaliste si on déroule la machine à gagner.

QUOI ? : POUR CHAQUE INFRASTRUCTURE DE PARTAGE DE DONNÉES, DÉROULER LA MACHINE À GAGNER

Les lignes directrices co-élaborées précisent considérablement le principe de plateforme publique présenté dans *Ubérisons l'État ! Avant que d'autres ne s'en chargent*. Elles passent 17 questions (voir la Figure 2) au crible des 7 infrastructures (voir la Figure 3) et en déduisent des grilles d'analyse, pour aider les porteurs d'infrastructures à capitaliser sur les enseignements existants. Elles constituent une première version qui a vocation à être adaptée et complétée.

**Pourquoi ? Une infrastructure de partage de données est-elle nécessaire ?
Quel doit être le cadre de valeurs associé ?**

1. [Besoins macro] Quels acteurs ont besoin de faire quoi d'un point de vue métier, et avec quelles données ?
2. [Besoins micro] Quels sont les premiers cas d'usage envisagés ?
3. Existe-il déjà des initiatives internes ou externes de construction d'une telle infrastructure ? De la part de quels acteurs ?
4. Quelles sont les raisons de l'échec des initiatives existantes et/ou les arguments contre la construction de l'infrastructure ?
5. Quel est l'échelon géographique pertinent (local, national, européen, international) ?
6. Quels sont les risques associés à cette infrastructure et quelles doivent être les valeurs qui l'encadrent ?
7. Si cette infrastructure n'existe/existait pas, quels sont/seraient les scénarios alternatifs et quels sont leurs risques ?

Quoi ? Quelles doivent-être les caractéristiques techniques, économiques et de gouvernance de cette infrastructure ?

8. Cette infrastructure doit-elle être unique ou peut-il en avoir plusieurs ?
9. Quelle doit être la gouvernance de cette infrastructure en terme de répartition des rôles public/externe ?
10. Quel doit être le modèle économique de cette infrastructure ?
11. Cette infrastructure doit-elle être centralisée ou décentralisée ?

Comment ? Quels sont les critères du développement et déploiement réussis de cette infrastructure ?

Transformation de l'interne :

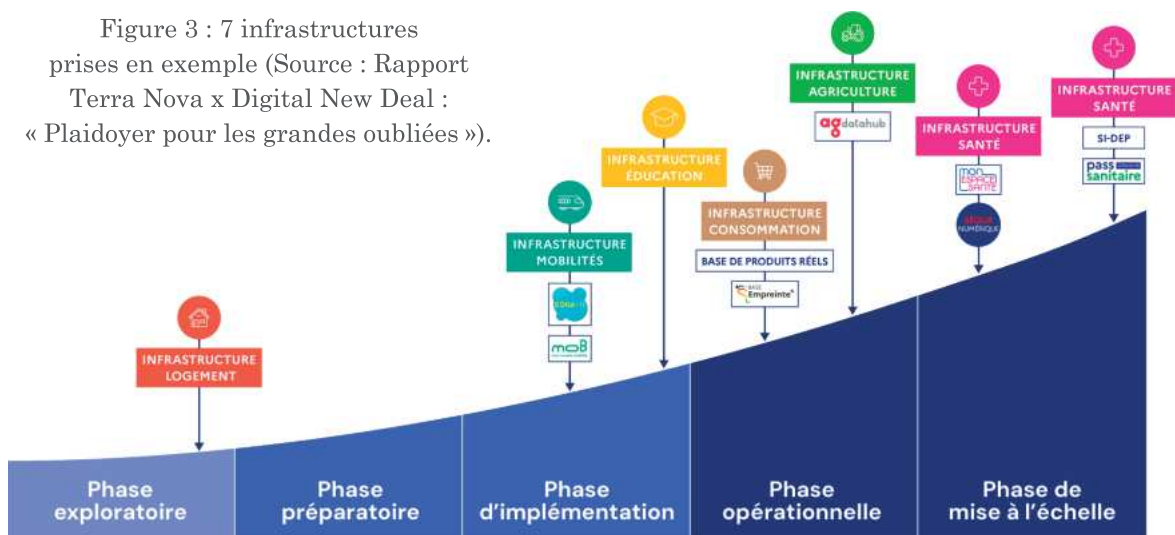
12. Quelle organisation et quel portage ?
13. Quels moyens humains et financiers ?
14. Quelle culture de travail et facteurs d'attractivité RH ?

Régulation de l'externe pour une co-construction et un déploiement efficace :

15. Quels leviers d'engagement des parties prenantes ?
16. Quels leviers coercitifs (« bâtons ») ?
17. Quels leviers incitatifs (« carottes ») ?

Figure 2 : Lignes directrices v0 (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

Figure 3 : 7 infrastructures prises en exemple (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).



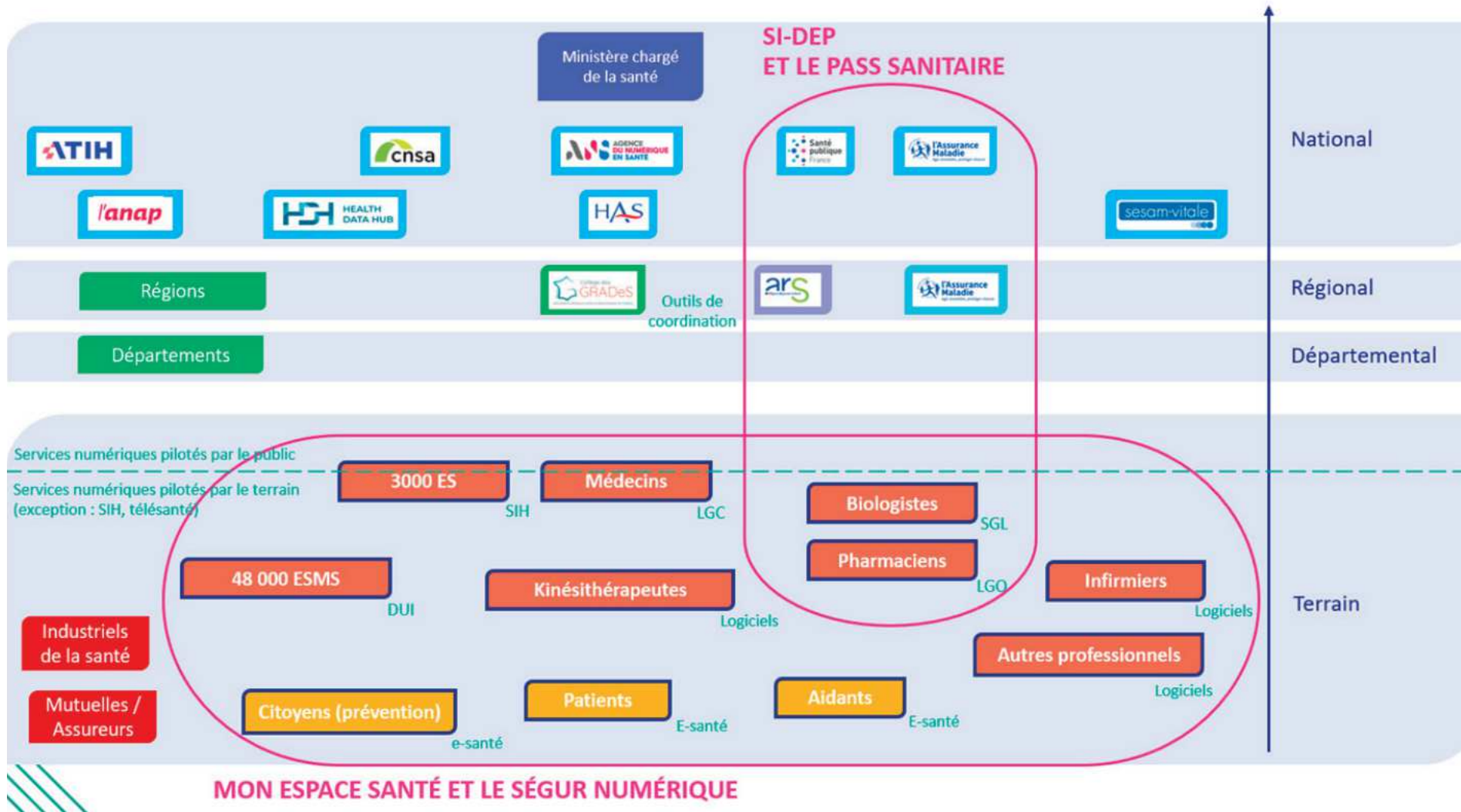


Figure 4 : Les acteurs de la santé (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

Chaque infrastructure est décrite schématiquement par : les acteurs concernés par le partage de données, la place de l'infrastructure dans le « bâtiment » de la politique publique concernée, ses cas d'usages macro et micro, et son architecture technique (voir exemples dans le secteur de la santé, cf. Figures 4, 5 et 6).



Figure 5 : Cartographie du secteur de la santé (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

Les lignes directrices aident d'abord les porteurs à élaborer de façon convaincante le « pourquoi ? », c'est-à-dire la raison d'être de l'infrastructure et le cadre de valeurs qui doit y être associée. Elles indiquent par exemple qu'il est nécessaire de présenter à la fois les besoins génériques que l'infrastructure permet de couvrir et des cas d'usage très précis qui aident les interlocuteurs à se projeter (voir exemples dans la Figure 7). Que les cas d'usage concernant l'utilisation des données à des fins primaires (par exemple permettre aux collectivités de réduire la consommation énergétique des écoles) sont moins polémiques et donc plus à même d'embarquer que ceux qui utilisent les données à des fins secondaires (par exemple connaître les écoles qui ont le plus de mal à remplacer les professeurs absents). Elles indiquent aussi que l'existence d'échecs précédents répétés est plus souvent un marqueur d'un vrai besoin que d'une fausse bonne idée (par exemple le DMP, prédécesseur de Mon Espace Santé, le « carnet d'information du logement »). Elles conseillent aussi d'explicitier la menace d'ubérisation : bien souvent, le *statu quo* est plus risqué que le changement. Cette partie est primordiale pour que toutes les parties prenantes soient absolument convaincues par le besoin et leur capacité à encadrer les risques, et ainsi ultra déterminées pour trouver des solutions aux obstacles qui ne manqueront pas de se présenter par la suite.

Si le besoin est confirmé, alors le « quoi ? » aide à arbitrer les questions qui se posent systématiquement sur les caractéristiques techniques, économiques et de gouvernance de

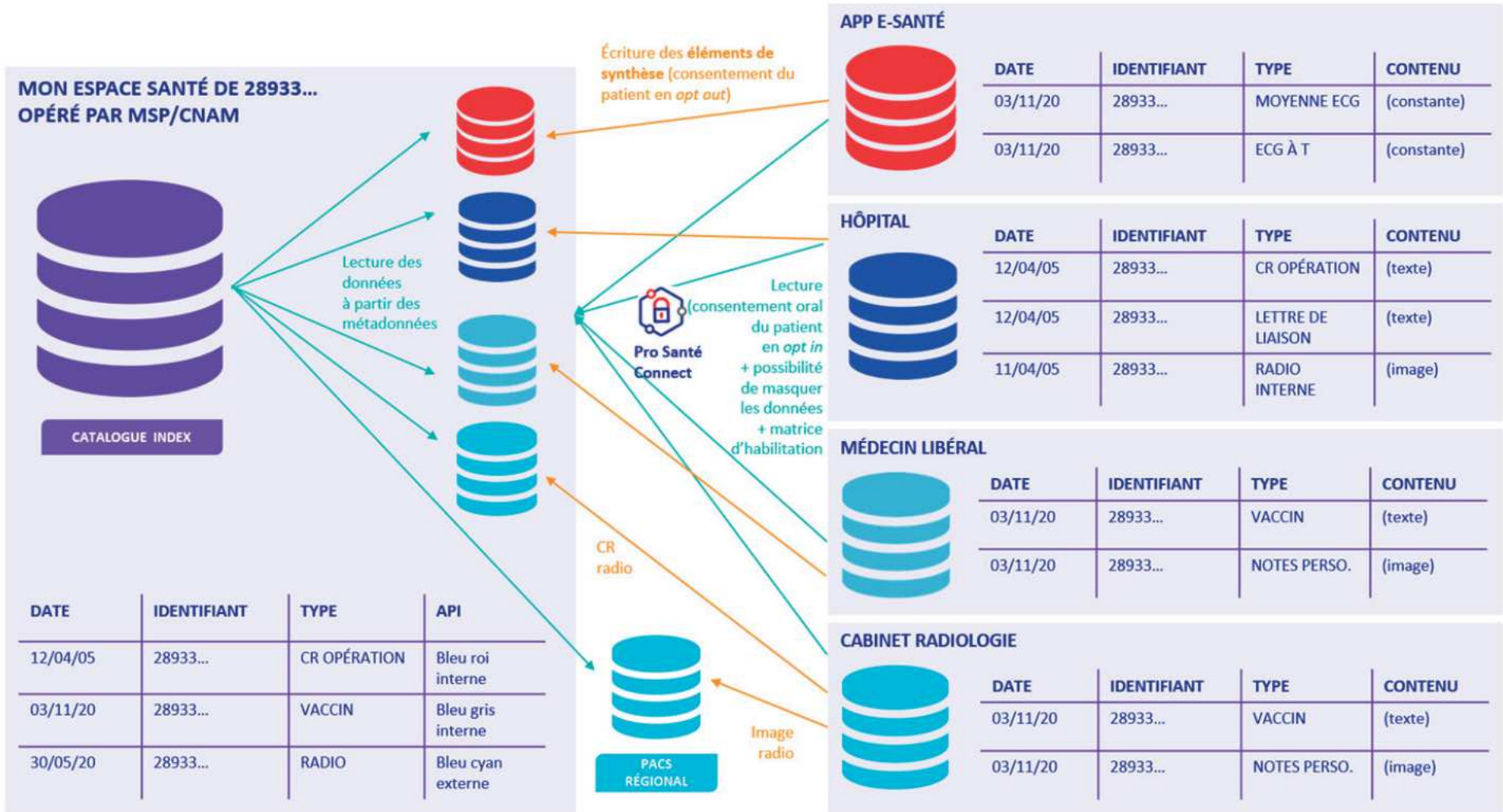


Figure 6 : L'architecture de Mon Espace Santé (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

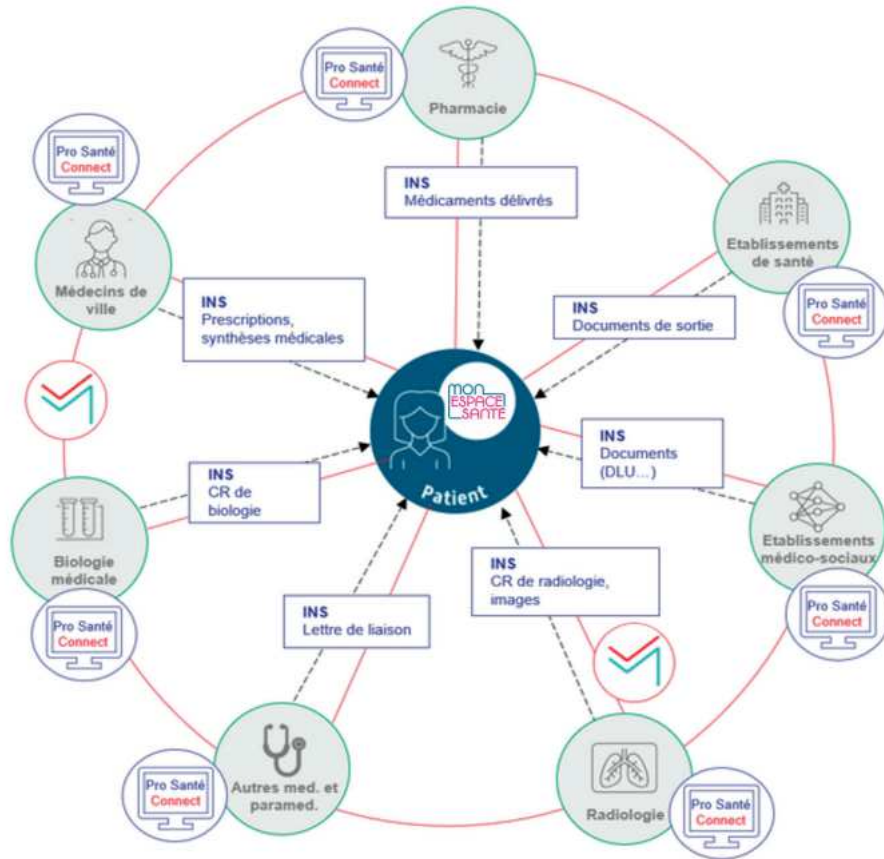


Figure 7 : Cas d'usage dans le secteur de la santé (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

[Macro] De façon générale :

- **Le patient** doit récupérer ses propres données personnelles pour être acteur de sa santé : comprendre et agir, demander un deuxième avis ...
- **Tous les professionnels sanitaires et médico-sociaux de l'équipe de soin** d'une personne (médecin traitant, infirmier, kiné, pharmacien, hôpital, EHPAD, SAMU, ...) doivent s'échanger des données la concernant pour bien le prendre en charge.
- **Les industriels de la e-santé** ont besoin d'accéder à certaines données des patients pour leur proposer, ainsi qu'au professionnels, des services numériques innovants (pour leur envoyer des rappels/alertes, services de prévention personnalisée, aide au diagnostic, ou entraîner une IA ...).

[Micro] Par exemple :

- Le patient récupère automatiquement son compte-rendu d'hospitalisation pour savoir comment l'opération s'est passée et montrer le document à son kiné pour qu'il adapte la rééducation en fonction.
- Le patient a perdu son ordonnance au moment où il doit retourner en pharmacie se faire dispenser des médicaments : il la retrouve dans son espace santé, et la transmet automatiquement ou par messagerie sécurisée au pharmacien en amont ou au moment de sa venue.
- Le patient a perdu sa preuve de vaccination (Covid, fièvre jaune ...) nécessaire à son voyage international. Il la retrouve dans Mon espace santé, où elle a été insérée automatiquement et où elle est stockée de façon sécurisée.
- Une application développée par le privé (ex: "comprendre mes résultats de biologie") peut, sans demander au patient de télécharger ses résultats mais juste en lui demandant l'autorisation d'accéder à ses compte-rendus de biologie structurés dans Mon espace santé, lui proposer un service qu'elle n'aurait pas pu lui proposer autrement.

l'infrastructure. Si cette infrastructure est indispensable à la mise en œuvre d'une politique publique, alors elle doit être unique et pilotée par le public sur le périmètre des cas d'usage concernés. C'est en ce sens que la feuille de route « Numérique et Données pour la planification écologique » a notamment annoncé le passage de la société Agdatahub sous gouvernance publique et la création d'un espace numérique du logement public. Il s'agit de fixer la frontière public/privé symbolisée par le sol du bâtiment : tout comme le carnet de santé papier n'est pas la propriété de Sanofi, le carnet de santé numérique ne peut être celle d'un acteur privé même français. En revanche, ce n'est pas à l'État de développer les logiciels de professionnels ou les applications de e-santé. Les questions suivantes sur le modèle économique et le caractère centralisé/décentralisé de l'infrastructure sont bien sûr importantes mais monopolisent souvent le débat au détriment des 15 autres questions des lignes directrices pourtant au moins aussi nécessaires.

Le « comment » dresse alors les 6 critères qui permettent de dérouler la machine à gagner. C'est de loin la partie la plus importante, la plus complexe et la plus fréquemment oubliée. C'est elle qui a permis de mettre sur pied SI-DEP en 3 semaines et de tenir la date du premier déconfinement alors que son prédécesseur « 3labos » peinait à voir le jour depuis 8 ans. Qui a fait de Mon espace santé, après seulement 2 ans d'existence, un outil déjà utilisé par près de 14 M de Français, qui collecte automatiquement les documents médicaux de 80 % des hôpitaux et de la majeure partie des professionnels de santé libéraux, alors que son prédécesseur le « DMP » restait une coquille vide depuis 2004. Les critères concernent à la fois la transformation interne des organisations publiques et leur façon de coconstruire et de réguler l'externe. En interne, il s'agit notamment de créer une équipe rattachée au plus haut niveau métier, qui consolide toutes les expertises numériques nécessaires. Cette équipe doit porter l'infrastructure en prenant le meilleur des deux mondes entre la culture *start-up* d'État et le modèle des grands projets informatiques. Cette méthode pourrait être résumée en : « Il faut aller des petits cailloux aux étoiles. Une étoile pour rêver et donner envie, et des petits cailloux pour se mettre en mouvement et baliser le chemin » (Jean-François Caron, sur la construction d'une ville durable). Cette partie donne aussi des conseils pour obtenir du portage politique ainsi que des moyens humains et financiers. En externe, il s'agit d'inverser la pratique où l'on décide parfois trop en chambre puis on met en œuvre de façon molle. Il convient de rentrer dans une démarche de démocratie participative rigoureuse, permanente et de terrain *via* les instances représentatives et les personnes en direct pour décider collectivement. Puis de mettre en œuvre efficacement ce qui aura été décidé ensemble *via* une sorte de dictature collaborative où, à la manière de Bruno Latour dans « Où atterrir ? », on établit la *to-do list* de chaque acteur et on s'assure à l'aide de moyens incitatifs et coercitifs que chacun aura les moyens de faire et fera effectivement, pour sortir du dilemme du prisonnier. L'ensemble du processus doit être ponctué de points d'étape ouverts à tous pour créer la confiance, permettre une coordination naturelle et se féliciter des étapes déjà franchies collectivement. En réalité, cette partie préconise simplement de remettre de la méthode et de l'humain pour rendre la démocratie plus plaisante et plus vivante.

COMMENT ? : ENDOSSER UNE VISION INTERMINISTÉRIELLE COMMUNE AU NIVEAU NATIONAL, EUROPÉEN ET INTERNATIONAL ET L'IMPLÉMENTER DE FAÇON ULTRA-DÉTERMINÉE

Comme pour tout sujet systémique, les infrastructures de partage de données posent des questions de gouvernance complexes. En interne, le sujet est aujourd'hui abordé par 3 types d'acteurs (voir la Figure 8). Les acteurs en charge du numérique « métier » lié aux politiques publiques (les agents des ministères de la Santé, de l'Éducation, de l'Agriculture, de l'Écologie, de l'Intérieur, de l'Économie, de la Défense, de la Justice, de la

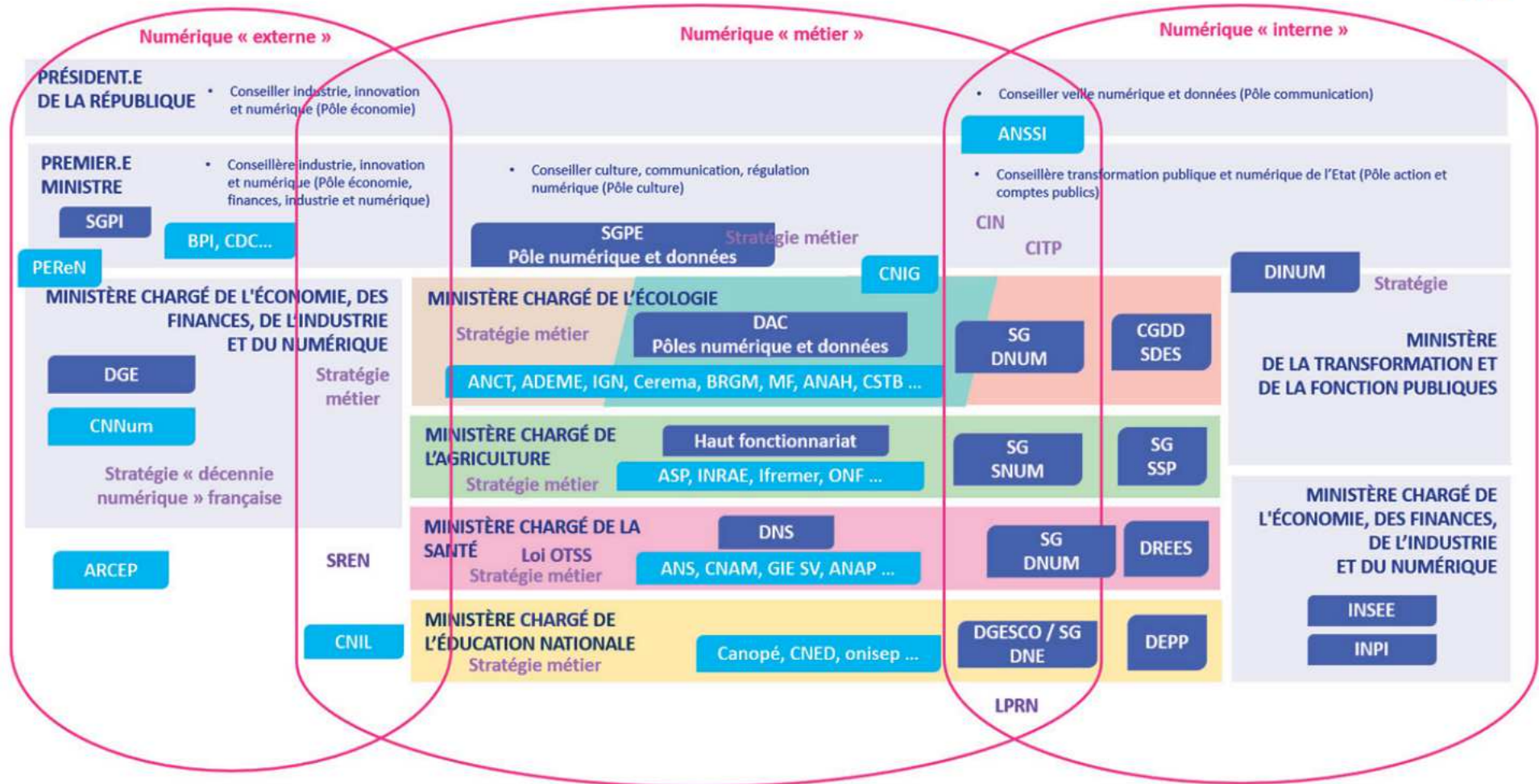


Figure 8 : La gouvernance actuelle au niveau national (Source : Rapport Terra Nova x Digital New Deal : « Plaidoyer pour les grandes oubliées »).

Culture... et les opérateurs qui les accompagnent) : ce sont eux qui doivent identifier les infrastructures nécessaires pour leurs politiques publiques et les mettre en œuvre. Les acteurs transversaux en charge du numérique « interne » lié à la transformation publique (les agents de la Dinum, de l'Insee...) : si historiquement ils s'occupent de l'outillage des agents et de leurs publications, leurs missions ont considérablement évolué vers des sujets métier avec l'*open data*, les *start-ups* d'État... Les acteurs transversaux en charge du numérique « externe » lié au développement économique (les agents de la DGE, du SGPI, de la Caisse des Dépôts, de la BPI...) : côté « bâtons », leur périmètre s'est élargi de la régulation des réseaux numériques physiques (télécom) à celles des plateformes et de données de plus en plus sectorielles en lien avec les régulations européennes ; côté « carottes », ils soutiennent la French Tech et plus globalement des entreprises qui ont un rôle à jouer dans les stratégies numériques des ministères.

Aujourd'hui, le manque de vision interministérielle commune entraîne parfois des voix dissonantes et des déficits de coordination et de portage politique qui nuisent à l'avancée des projets au niveau national et nous empêchent de jouer pleinement un rôle pionnier au niveau européen et international. Ces lignes directrices initiales pourraient être challengées et enrichies avec les retours de tous les acteurs internes et externes pertinents pour donner lieu à une première version endossée politiquement. La responsabilité d'une infrastructure, dès lors qu'elle est sectorielle, doit être portée par le ministère concerné pour qu'elle soit construite au plus proche des enjeux de terrain. Il doit être capable d'expliquer de façon convaincante et pédagogique le besoin à toutes les parties prenantes. Il serait utile que la Dinum et la DGE forment quant à elles un binôme « interne / externe » qui joue à la fois le rôle de « gendarme », pour s'assurer du respect de la répartition des rôles publics/privés et plus largement des lignes directrices, et de « soutien » en aidant les ministères à négocier la gouvernance et les moyens humains et financiers, en mettant à disposition des expertises pointues, etc.

Un point fait largement consensus : au-delà de la méthode, au bout du bout, tout est une histoire de personnes. D'agents publics en administration centrale, en administration territoriale, chez les opérateurs, et d'acteurs externes courageux qui se battent au service d'une mission commune. Qui ont cette capacité à surmonter la bureaucratie, à braver le construit, à proposer de nouveaux cadres et processus mieux à même de servir l'intérêt général. Qui « trahissent » au sens d'Aurélien Barrau dans *L'Hypothèse K*, à propos des scientifiques qui questionnent le rôle de la science et le processus scientifique. Qui trahissent « non pas les paroles données, encore moins l'honnêteté. Il s'agit de trahir les pratiques hérités et les suivismes implicites. Trahir l'origine. Trahir l'attendu. Trahir l'inertie. De déplaire par droiture, de décevoir par intégrité. Puisque ne pas trahir, face au dévoiement, c'est trahir plus. De trahir par amour, [...] dans une infidélité authentique. Pour prendre le temps d'affronter les injonctions contradictoires. Pour accueillir un peu de la possibilité d'une méta fidélité à la vie. À l'avenir, à l'improbable, à la beauté, à l'espoir ». Le présent travail est aussi l'occasion de leur dire merci.

BIBLIOGRAPHIE

DIGITAL NEW DEAL & TERRA NOVA (2024), « Plaidoyer pour les grandes oubliées : les infrastructures publiques de partage de données », rapport collaboratif coordonné par Laura Létourneau, <https://www.thedigitalnewdeal.org/plaidoyer-pour-les-grandes-oubliees-les-infrastructures-publiques-de-partage-de-donnees/>

Les systèmes de santé

Par Michel BARTH

Président et co-fondateur d'ENoving

Le numérique investit et bouleverse le secteur de la santé et ce mouvement n'est pas près de ralentir, à la faveur notamment du développement de l'internet haut débit, du *cloud*, de la mobilité et des objets connectés (*Internet of Things* ou IoT). Au-delà d'une évidente rationalisation et efficience des soins permise par la numérisation des organisations et processus, la numérisation de la santé apporte aussi et surtout une promesse qualitative concernant l'augmentation des chances des patients d'être diagnostiqués de manière juste le plus tôt possible, puis d'être soignés et rééduqués efficacement, tout en fluidifiant les épisodes de soins notamment pour les hospitalisations ainsi que pour les pathologies longues ou chroniques. Toutes les données générées transitent sur des réseaux fixes et mobiles pour les usages les plus divers au service du parcours de soins du patient. Il en résulte un besoin toujours croissant d'infrastructures numériques fixes et mobiles pour soutenir ces flux.

L'ACCÉLÉRATION DE LA NUMÉRISATION DE LA SANTÉ

La révolution numérique en santé

Le numérique investit et bouleverse le secteur de la santé comme partout ailleurs et ce mouvement n'est pas près de ralentir, à la faveur notamment du développement de l'internet haut débit, du *cloud*, de la mobilité, des objets connectés (*Internet of Things* ou IoT), mais aussi d'autres technologies telles que le métavers ou la robotique. Les Dispositifs Médicaux (DM) se digitalisent et l'on voit même apparaître une nouvelle catégorie dite de Dispositifs Médicaux Numériques (DMN)¹.

C'est ainsi que des données de plus en plus nombreuses sont générées, manipulées par des logiciels et DM, offrant par ailleurs des perspectives majeures d'usages secondaires. Selon un rapport de l'Assemblée nationale sur le dossier médical partagé et les données de santé², 153 exaoctets³ de données de santé ont été produits dans le monde en 2013 à comparer aux 2 314 exaoctets qui ont été produits en 2020. Ces données proviennent à la fois des dossiers médicaux des patients, de l'imagerie médicale, de résultats de tests biologiques, de données génétiques parfois ou encore de données de capteurs physiologiques qui peuvent délivrer leurs informations en continu, sans compter les informations fournies par les patients eux-mêmes.

¹ <https://esante.gouv.fr/produits-services/dispositifs-medicaux-numeriques>

² <https://www.vie-publique.fr/eclairage/289281-donnees-numeriques-de-sante-quels-enjeux-pour-quel-progres-medical#:~:text=En%202013%2C%20selon%20un%20rapport,314%20exaoctets%20ont%20été%20produits>




³ 1 exaoctet = 1018 octets.

Parmi ces données, l'imagerie médicale tient une place particulière et historique⁴, l'ensemble des modalités d'imagerie (Scanner, IRM, PET Scan, radiologie conventionnelle, mammographie, échographie...) ayant progressivement évolué vers les technologies numériques, la dernière évolution en date étant la numérisation des lames d'anatomocytopathologie, ceci à la faveur de l'émergence très tôt de standards de codage tels que DICOM⁵. De fait, la nouvelle imagerie médicale produit un volume de données de plus en plus important, 80 millions d'actes d'imagerie étant réalisés en France chaque année⁶, avec des images dont la taille varie entre quelques mégabits à plusieurs gigabits, permettant une exploitation élaborée (reconstruction 3D, imagerie vasculaire, imagerie moléculaire), une optimisation des temps d'examen et des examens moins invasifs.

Les types d'usage et niveaux d'impact du numérique

Bien que toute forme de classification soit par essence difficile, il est possible de distinguer trois grandes typologies d'usage du numérique en santé :

- en premier lieu, le numérique au service de l'organisation des soins, avec notamment les logiciels utilisés pour la gestion des établissements sanitaires, comme les Systèmes d'Information Hospitaliers (SIH), les applications utilisées pour l'échange et le partage des données entre professionnels voire avec les patients, ainsi que les systèmes de santé publique permettant la prise de décision au service de la collectivité ;
- en second lieu, le numérique dans l'acte de soins, intégré notamment dans des DM ou constituant en lui-même un DMN, transformant la pratique même des professionnels de santé ou le vécu des patients, avec un ensemble hétérogène d'usages les plus variés, au service du diagnostic, de la chirurgie, de la rééducation, de l'information

Niveau Systémique	<p>Crée un nouvel écosystème et de nouvelles relations entre acteurs</p> <p>Apporte qualité et réduction de coût</p> <p>Peut radicalement changer la donne</p> <p>Nécessite un programme d'innovation dédié</p>	<p>Ex. Hôpital du Futur</p> 
Niveau Processus	<p>Impacte la répartition des tâches sur un processus à l'intérieur d'une entreprise ou impliquant plusieurs structures</p> <p>Apporte qualité et gain de temps</p> <p>Nécessite de repenser les processus et requiert une conduite du changement</p>	<p>Ex. Télémédecine</p> 
Niveau Poste de travail	<p>Impacte un professionnel ou une petite équipe autour d'une tâche isolée</p> <p>Apporte qualité et productivité</p> <p>Nécessite une formation</p>	<p>Ex. Stéthoscope numérique</p> 

Source : Michel BARTH

Source des 3 illustrations : schémas générés par Chat GPT 4

Figure 1 : Les niveaux d'impact du numérique en santé (Source : Michel Barth).

⁴ <https://www.lesechos.fr/idees-debats/editos-analyses/les-nouvelles-dimensions-de-limagerie-medicale-1153831>

⁵ Digital imaging and communications in medicine.

⁶ <https://fnmr.fr/wp-content/uploads/2022/01/FNMR-Livre-Blanc-de-limagerie-medicale-en-France-1.pdf>

et de l'éducation du patient, de son bien-être, de la réduction des handicaps, de la prévention, etc. ;

- enfin, le numérique au service de l'enseignement et de la recherche.

Sur un autre plan, il est possible de classer les niveaux d'impact du numérique en santé en trois catégories d'impact croissant représentatives de trois paliers de complexité et de natures différentes d'échanges de données.

Le besoin qui en résulte en matière d'infrastructures numériques

Toutes les données générées transitent sur des réseaux fixes et mobiles, soit qu'elles aient besoin d'être sauvegardées dans le *cloud* ou bien d'être échangées entre professionnels de santé pour les besoins les plus divers au service du parcours de soins du patient. Il en résulte un besoin toujours croissant d'infrastructures numériques fixes et mobiles pour soutenir ces flux.

Ainsi, par exemple, les GHT (Groupements Hospitaliers de Territoire), constitués en général de plusieurs établissements distants sur chaque territoire, ont besoin de disposer de réseaux sécurisés de type VPN (*Virtual Private Network*) permettant des échanges à haut débit à l'échelle territoriale s'appuyant notamment sur des réseaux fixes de type fibres optiques.

Les logiciels métiers utilisés par les professionnels reposent de plus en plus sur une architecture *cloud* nécessitant un accès permanent à un réseau de débit suffisant, avec un recours à des ressources de calcul distant en temps réel, par exemple en lien avec l'usage de moteurs d'Intelligence Artificielle (IA).

Dans le domaine des images numériques, celles-ci peuvent être stockées et indexées notamment dans des PACS (*Picture Archiving and Communication System*) qui sont des dispositifs matériel et logiciel qui permettent de traiter et d'interpréter des images médicales numériques. Il y a alors besoin d'accéder en permanence à distance à ces PACS au travers de réseaux haut débit permettant un accès rapide à l'information.

Dans les territoires, les professionnels de santé libéraux ont de plus en plus besoin d'échanger de l'information à tout moment de leur exercice entre eux et avec les patients, avec un besoin de mobilité accrue par exemple concernant les infirmières libérales à domicile.

La télémédecine, qui se développe considérablement depuis la crise de Covid-19, nécessite une mise en relation distante entre le patient et un professionnel de santé, avec un besoin toujours croissant de données échangées pour la circonstance (vidéos, données physiologiques, etc.), requérant une large infrastructure couvrant le territoire.

Les services d'urgence ont besoin sur le terrain en mobilité d'accéder le plus rapidement possible à une information de qualité concernant les blessés qu'ils doivent secourir.

Tout ceci se fait dans un contexte sécuritaire fort, les données personnelles de santé constituant un matériau particulièrement sensible très protégé sur le plan réglementaire, une brèche de sécurité dans un hôpital pouvant stopper son fonctionnement pendant plusieurs jours. C'est ainsi que les flux de données transitant sur les réseaux fixes et mobiles doivent donc être particulièrement protégés.

Les usages engendrés

De nouveaux usages apparaissent qui changent les frontières entre les acteurs ainsi que le paysage du parcours de soins du patient. Ces usages sont multiples et impactent en particulier le vécu des professionnels de santé, avec une vision transformée du rôle de

l'hôpital et de la médecine de proximité dans les territoires. Ces nouveaux usages transforment également le vécu des patients, pour mieux vivre avec la maladie et mieux vieillir demain.

Les exemples sont déjà nombreux qui témoignent de ces transformations. Il est ainsi possible aujourd'hui, par exemple, dans des registres très différents de : travailler facilement en équipe de soins pluridisciplinaire sur un territoire autour de cas patients complexes par des plateformes numériques d'échange et de partage sécurisées dédiées santé ; affiner les diagnostics et traitements pour des maladies telles que le cancer par l'intelligence artificielle et les jumeaux numériques (jumeau numérique du patient) ; permettre une sortie d'hôpital bien plus rapide qu'auparavant par une télésurveillance à domicile post-opératoire ; intervenir à distance pour un professionnel de santé par la téléconsultation voire par la téléchirurgie, la télémanipulation ou la télééducation ; faciliter l'enseignement de la chirurgie par la réalité virtuelle ; faciliter et accélérer le développement de nouvelles molécules par des tests sur cohortes de patients simulés ; etc.

Ces usages nouveaux facilitent l'émergence de structures nouvelles comme les Communautés Professionnelles Territoriales de Santé (CPTS), qui sont en plein essor depuis la Loi de modernisation de la santé du 26 janvier 2016 et le programme « Ma Santé 2022 » avec 754 structures recensées par le ministère de la Santé⁷. Ces communautés permettent à la médecine de ville et aux professionnels de santé libéraux de s'organiser sur les territoires avec l'aide du numérique pour travailler mieux ensemble au service des parcours de soins patients, notamment pour les cas les plus complexes (pathologies chroniques ou patients âgés).

LES PROMESSES DE LA NUMÉRISATION DE LA SANTÉ

Au-delà d'une évidente rationalisation et efficience des soins permise par la numérisation des organisations et processus, la numérisation de la santé apporte aussi et surtout une promesse qualitative concernant l'augmentation des chances des patients d'être diagnostiqués de manière juste le plus tôt possible, puis d'être soignés et rééduqués efficacement, tout en fluidifiant les épisodes de soins notamment pour les hospitalisations ainsi que pour les pathologies longues ou chroniques.

Nous proposons ci-après de mettre en évidence deux promesses particulièrement emblématiques pour l'avenir de notre système de santé, relatives, pour l'une, à l'intelligence artificielle en santé et, pour l'autre, à l'humain réparé et augmenté.

Les promesses concernant l'Intelligence Artificielle (IA) en santé

Dans un monde où la technologie évolue à un rythme exponentiel, l'IA se distingue comme un catalyseur de changement, particulièrement dans le secteur de la santé. L'analyse de données massive est une pratique de longue date en santé avec la génomique en particulier, et le *big data* a tenu ses promesses en imagerie médicale notamment. Avec la disponibilité croissante de données de toutes sortes, l'IA s'est récemment introduite dans tous les domaines de la santé au travers de multiples applications spécialisées et en s'insérant dans la plupart des dispositifs médicaux, par exemple pour détecter précocement des pathologies, formuler des diagnostics plus précis, personnaliser des traitements ou automatiser des tâches répétitives. Depuis peu, l'IA générative ouvre de nouvelles perspectives encore plus disruptives pour le système de santé, par exemple, pour l'aide

⁷ <https://sante.gouv.fr/systeme-de-sante/structures-de-soins/les-communautés-professionnelles-territoriales-de-sante-cpts/>

au diagnostic, la rédaction de comptes rendus médicaux et d'autres applications plus inattendues encore. Ceci n'est pas sans poser de nombreuses questions éthiques et réglementaires, sans compter la nécessité de préparer et nettoyer les données.

De nombreuses publications récentes décrivent cette situation et traitent de l'apport de l'IA en santé et son adoption croissante⁸. Nous avons dans ce contexte récemment coordonné la sortie d'un hors-série de la revue *Télécom*⁹ dans lequel nous avons voulu explorer au travers de plusieurs articles de multiples facettes de l'IA appliquée à la santé pour des usages variés, sans prétendre à l'exhaustivité, et en cherchant à mettre en lumière les défis posés et les perspectives engendrées. En particulier, nous y abordons plusieurs exemples d'usage spécifiques et variés concernant l'IA en imagerie médicale (P.H. Conze, IMT & LaTIM, 2024), l'IA en radiothérapie (C. Robert, Université Paris-Saclay & IGR, 2024), l'IA pour l'endoscopie (Pr X. Dray, Hôpital Saint Antoine & Augmented Endoscopy, 2024), l'IA et la prescription médicamenteuse (V. Bouvier, Vidal Group, 2024), ou enfin l'IA en assurance santé (V. Lacam-Denoel, Proxicare, 2024). Nous y abordons également des usages spécifiques de l'IA générative concernant la génération de protéines (J. Holland, journaliste, 2024). Nous traitons enfin le sujet de la technique de l'apprentissage fédéré et ses avantages (M. Clerc avec A. Bellet et M. Lorenzi, Inria, 2024), concernant le problème de la confidentialité des sources de données en santé. Dans ce modèle, un ensemble d'acteurs se coordonne pour entraîner des algorithmes d'IA à partir de l'ensemble de leurs jeux de données, sans que ces données soient à aucun moment partagées avec un tiers. Cette technique, qui pose encore des défis, est prometteuse dans le domaine de la santé, où les données sensibles sont difficiles à partager.

Tout ceci témoigne que les patients que nous sommes peuvent s'attendre à l'avenir à un afflux d'outils basés sur l'IA dans toutes les facettes de notre santé avec des inquiétudes et questions légitimes que cela peut parfois susciter.

Les promesses concernant l'humain réparé et augmenté

La question de l'humain réparé ou augmenté suscite de nombreux rêves depuis plusieurs années tout en posant des problèmes éthiques épineux. Sans s'aventurer sur le terrain du transhumanisme, nous proposons de nous focaliser sur plusieurs exemples de ce qui peut être apporté à l'être humain dans les années à venir, dans les domaines de la santé et du handicap.

Les progrès concernant l'humain réparé sont constants en médecine. Les techniques médicales combinées aux biotechnologies connaissent déjà et vont continuer à connaître des innovations radicales. Les technologies robotiques, électroniques et numériques jouent un rôle crucial pour venir soutenir certaines de ces innovations médicales souvent en lien avec la réparation d'un handicap, en témoignent de nombreux exemples : impression 3D d'organes ou de tissus humains, usage de nanotechnologies pour la vision, main artificielle connectée au cerveau et commandée par la pensée, interfaces neuronales, organes artificiels (cœur, pancréas, etc.), prothèses intelligentes, exosquelettes pour les personnes hémiplegiques ou amputées, implant dans la colonne vertébrale permettant à une personne paralysée de remarcher, etc. À noter que les technologies en question voient en parallèle bien d'autres usages dans le monde de la santé hors le sujet de l'humain réparé. C'est par exemple le cas de l'impression 3D qui peut aussi servir à fabriquer des objets et dispositifs médicaux divers au service des soins.

⁸ Voir la liste de ces publications à la fin de la bibliographie.

⁹ <https://www.telecom-paris-alumni.fr/fr/revue/numeros/ia-et-sante-numerique-et-defense/3577>

En allant plus loin, à la faveur de la réparation d'un handicap ou d'un organe, les nouvelles technologies pourraient en même temps s'aventurer sur le terrain de l'humain augmenté. En effet, rien n'interdit pour les technologies d'aller plus loin que simplement palier une carence/impotence fonctionnelle et aller vers une augmentation des fonctions humaines (plus de force, meilleure vision, etc.), ceci de facto pouvant poser des questions éthiques complexes. C'est par exemple le cas de récentes avancées de chercheurs chinois travaillant sur un utérus artificiel géré par IA. Par ailleurs, ces nouvelles technologies peuvent déborder du domaine de la santé pour des applications dans le domaine du service, de l'industrie ou du militaire concernant notamment les exosquelettes.

LE CADRE DONNÉ PAR LES POUVOIRS PUBLICS

Au-delà d'une réglementation européenne qui se renforce de manière générale sur le numérique (RGPD, IA Act, marquage CE, etc.) et plus spécifiquement dans le domaine de la santé (Dispositifs Médicaux ou DM notamment), la France s'est dotée depuis quelques années d'un programme ambitieux d'accélération de la santé numérique. C'est ainsi que les pouvoirs publics ont mis sur pied une feuille de route du numérique en santé¹⁰ qui a pour objectif d'établir le cadre d'urbanisation et de décrire les directives réglementaires aux acteurs impliqués dans la création, le développement et la maintenance de services numériques en santé, formant ainsi la maison du numérique en santé qui symbolise ce cadre et incarne la vision d'État plateforme.



Figure 2 : La maison du numérique en santé
(Source : Feuille de Route du Numérique en Santé 2023-2027).

¹⁰ <https://gnius.esante.gouv.fr/fr/programmes-nationaux/feuille-de-route-du-numerique-en-sante>

Dans ce contexte, l'État a mis en place un programme Ségur Numérique de référencement et financement de solutions numériques interopérables qui s'impose à tous. De nombreuses initiatives de subvention publiques pour favoriser le développement de l'innovation ont été lancées, comme l'article 51 qui permet d'expérimenter des solutions organisationnelles nouvelles, ou les tiers lieux d'expérimentation qui permettent aux solutions numériques innovantes de trouver leurs lieux d'usage. Un écosystème d'acteurs institutionnels s'est mis en place (Agence du Numérique en Santé ou ANS, *Health Data Hub* ou HDH, Agence de l'Innovation en Santé ou AIS, Paris Santé Campus, G_Nius, etc.). Le cadre de la télémédecine s'est considérablement étoffé pour faciliter les téléconsultations notamment et permettre l'émergence de services de télésurveillance médicale remboursés par la sécurité sociale, faisant de la France un pays pionnier en la matière. Des services socles nationaux ont été déployés, avec en premier lieu Mon Espace Santé, mais aussi les Messageries Sécurisées de Santé (MSSanté), les solutions d'échange et partage (*e-parcours*), les plateformes de Services d'Accès aux Soins (SAS) ou encore la *e-prescription* qui se généralise. En relais de ces services, un ensemble de plateformes régionales d'échange et de partage de documents et imagerie médicale se développe sous l'impulsion des Agences Régionales de Santé et leurs GRADeS¹¹, comme PREDICE dans les Hauts-de-France ou PARCEO dans le Grand-Est.

Cet environnement a de nombreux avantages :

- il accompagne le développement du marché des industriels du numérique en santé, en sécurisant leur développement dans la mesure où ceux-ci bâtissent sur les autoroutes de l'État plateforme ;
- il assure une vision plus urbanisée et pérenne de l'usage des solutions numériques pour les établissements et professionnels sanitaires, avec une meilleure interopérabilité entre solutions, facilitant notamment le fonctionnement des structures hospitalières ;
- il permet aux citoyens de mieux avoir accès au système de soins et piloter leur santé en tout point du territoire.

L'AVENIR DE LA MOBILITÉ EN SANTÉ

Parmi les technologies d'infrastructure numérique pouvant avoir un impact particulier sur la santé à l'avenir, nous proposons de mettre l'accent sur le développement des réseaux 5G et autres réseaux mobiles du futur (5G+ avec la 6G et autres), ceci à la faveur de travaux de réflexion menés avec l'Institut Mines Télécom et plusieurs partenaires en 2021 en anticipation d'un appel à projets sur la souveraineté dans les réseaux de télécommunication en lien avec la 5G.

Perspectives offertes par les réseaux mobiles du futur

Les réseaux 5G+ peuvent amener une véritable rupture dans l'offre de soins tout au long du parcours du patient. La 5G promet en effet plusieurs améliorations par rapport aux technologies mobiles précédentes, parmi lesquelles une plus faible latence, une vitesse de transfert des données extrêmement élevée, et un accès massif des objets connectés (contexte IoT). La 5G autorise également une plus grande capacité de découpage du réseau en plusieurs tranches virtuelles (*slices*), exploitées chacune depuis une infrastructure partagée commune, mais adaptées individuellement aux cas d'usages et services ciblés (par exemple besoins particuliers en fiabilité, latence, bande passante, couverture réseau, etc.). De la même manière qu'une entreprise peut aujourd'hui déployer son propre réseau privé, l'équivalent est possible en 5G.

¹¹ Groupement Régional d'Appui au Développement de la *e-Santé*.

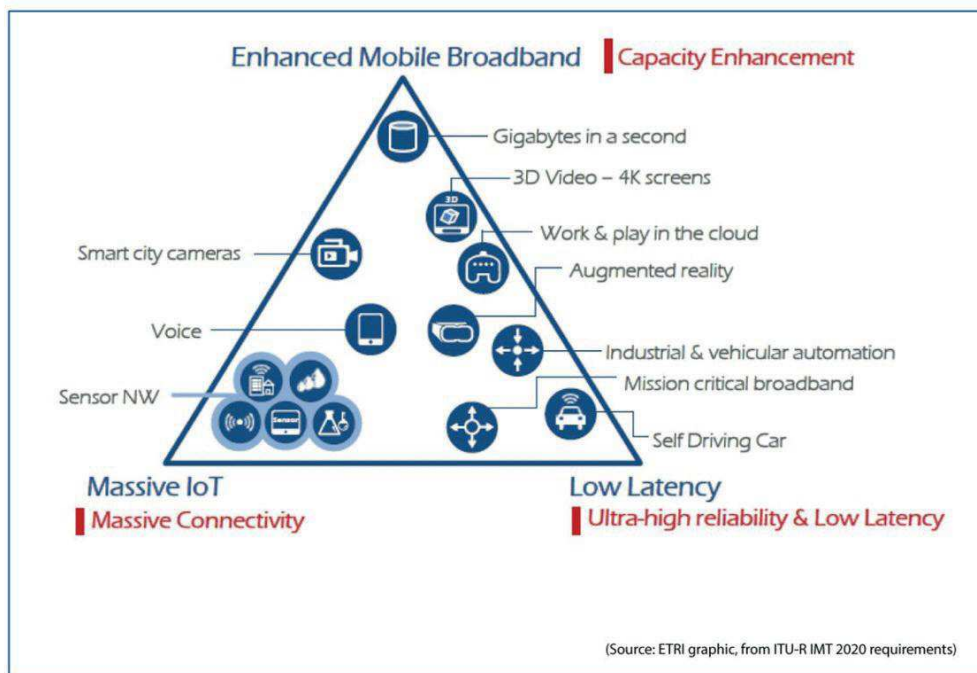


Figure 3 : Haut débit mobile amélioré (Source : ETRI graphic).

Ces caractéristiques contribueront à accélérer la transformation digitale de nombreux secteurs dont la santé. En effet, selon Gartner, le domaine de la santé est l'un des premiers à avoir capté l'attention de l'industrie des télécoms et constituera 14 % des cas d'usages liés à la 5G¹². Au niveau mondial, la base installée d'équipements de santé connectés en 5G devrait atteindre 3,2 millions d'unités en 2028 (contre 50 000 en 2020, soit un taux de croissance annuel moyen de 68 %).

Cas d'usage associés

Dans un tel contexte, les cas d'usages liés à la 5G en santé pourraient être classés selon deux grandes catégories :

- Des cas d'usages de masse : dans une certaine mesure, ceux-ci sont déjà réalisables aujourd'hui sans 5G mais leur usage sera amplifié grâce à l'amélioration du débit, de l'infrastructure IoT, et potentiellement du découpage du réseau. Ils permettent d'améliorer l'autonomisation du patient, la télémédecine et les communications entre de nombreux services (en interne à l'hôpital ou entre plusieurs institutions par exemple). D'après Gartner, la collecte de données cliniques, les outils de dépistage et diagnostic ainsi que la télémédecine mobile constitueront 80 % des usages de la 5G en santé.
- Des cas d'usages critiques spécifiques, rendus possibles par la 5G, et qui bénéficient en premier lieu de sa faible latence afin de connecter des objets et ressources critiques en temps réel, comme la chirurgie robotique à distance ou la formation *via* la réalité virtuelle /augmentée ou encore l'ambulance connectée (voir plus bas).

La Figure 4 (*cf.* page suivante) illustre les degrés d'adoption de différents cas d'utilisation en fonction de la pertinence de la 5G.

¹² Market Trends: 5G for Healthcare, Gartner 2020.

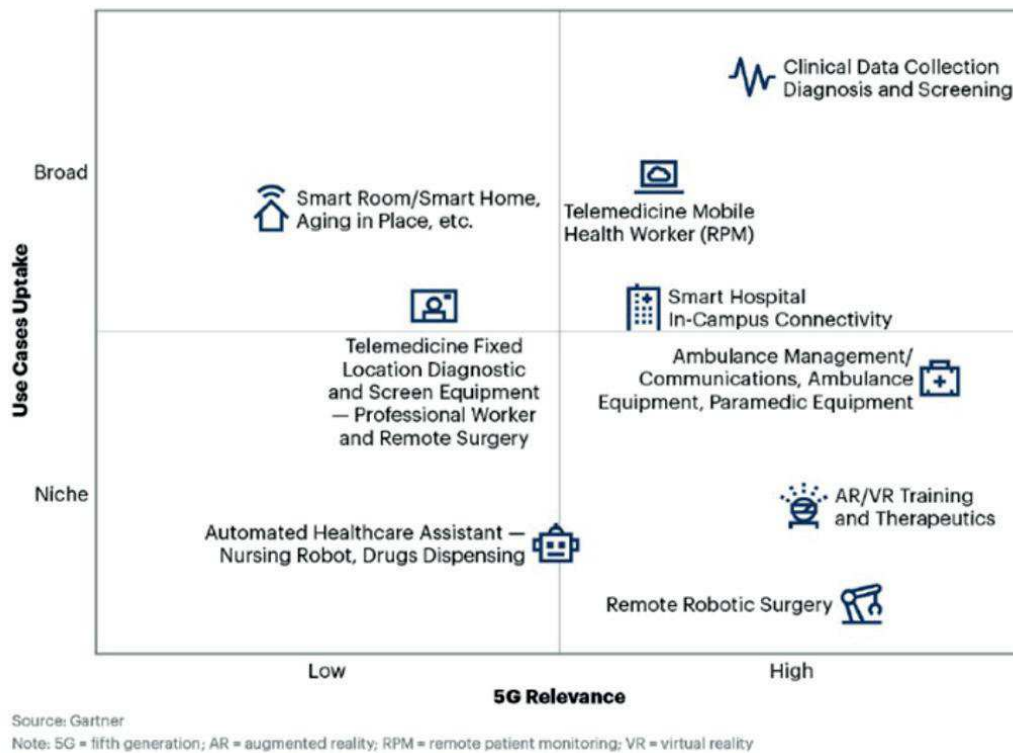


Figure 4 : Cas d'utilisation de la 5G d'ici 2025 (niveau d'adoption vs pertinence de la 5G) (Source : Gartner 2020).

Focus sur un cas d'usage emblématique : l'ambulance connectée du futur¹³

L'ambulance 2.0, ou ambulance connectée à la 5G, pourrait faire progresser la prise en charge du patient dans un contexte d'urgence entre le moment de l'appel conduisant à activer l'ambulance et son arrivée à l'hôpital, le terme « ambulance » étant ici à comprendre comme véhicule utilisé par les services d'urgence. Les équipes et véhicules de secours, dont la priorité est de réduire les délais de prise en charge des patients en situation d'urgence, auront à leur disposition des dispositifs embarqués qui permettront à la fois d'optimiser ces délais de prise en charge, dont les temps de transport, et d'améliorer les capacités diagnostiques et thérapeutiques à travers une meilleure communication avec les ressources hospitalières (services de soins critiques, plateaux techniques, etc.).

Pour ce faire, l'ambulance connectée doit combiner de nombreuses technologies avancées pour permettre d'améliorer l'efficacité, la sécurité et la qualité de prise en charge pré-hospitalière des patients souffrant de pathologies urgentes et dont le pronostic est le plus défavorablement associé au retard de soins (arrêt cardiaque, accident vasculaire cérébral, traumatisme grave, infarctus du myocarde, etc.).

À titre d'exemple, selon les experts ayant travaillé avec nous sur le sujet en 2021, en cas d'arrêt cardiaque, chaque minute passée entre l'effondrement et le début du massage, la défibrillation ou l'introduction d'adrénaline, entraîne une diminution de chance de survie d'environ 10 %. En cas de traumatisme grave, chaque tranche de 10 minutes supplé-

¹³ <https://www.usine-digitale.fr/article/ambulance-connectee-retransmission-video-en-4k-vr-depuis-amsterdam-juniper-networks-incite-a-faire-le-succes-de-la-5g-en-imaginant-les-cas-d-usage-N856785> ; https://www.sfm.u.org/fr/actualites/actualites-de-l-urgences/le-chu-de-rennes-experimente-deux-cas-d-usages-de-la-5g/new_id/67749

mentaires entre la prise en charge de l'appel et jusqu'à l'arrivée à l'hôpital est également associée à une réduction de survie de 10 %. Dans les accidents vasculaires cérébraux, chaque minute supplémentaire depuis le début des symptômes jusqu'au traitement entraîne la mort de 2 millions de neurones (réduction de 26 % de chance de bon pronostic pour chaque tranche de 30 minutes supplémentaires).

En conséquence, les capacités offertes par l'arrivée de la technologie 5G pourraient faire de l'ambulance connectée un élément central dans le parcours de soins urgents et non programmés. Elle constituerait une sorte de dispositif mobile intelligent, faisant évoluer la fonction naturelle du transport du patient vers des fonctions agrégées à haute valeur ajoutée au service des équipes de secours médical.

CONCLUSION GÉNÉRALE

Les nouveaux outils apportés par le numérique et les données transportées sur les autoroutes de l'information nous permettent d'aborder avec espoir les défis posés au système de santé français, et notamment le vieillissement de la population et la recrudescence des maladies chroniques, dans un contexte de développement exponentiel et de fragmentation des savoirs, avec une organisation du système de santé problématique, cloisonnée, où existent de véritables déserts médicaux, et faisant face à un déficit chronique de ressources.

BIBLIOGRAPHIE

RAPPORT DE LA COUR DES COMPTES (SÉCURITÉ SOCIALE 2024), « Le système national des données de santé : un vaste gisement d'informations à mieux exploiter », chapitre XI, mai 2024.

RAPPORT DU SÉNAT (2023), « Données de santé : une réforme encore en cours de chargement », n°873, 12 juillet 2023.

BULLETIN DE L'ORDRE NATIONAL DES MÉDECINS, Numéro Spécial, « Santé : la révolution numérique », janvier 2022.

LIVRE BLANC « 5G ET SANTÉ », Comité Stratégique de Filière (CSF) Infrastructures Numériques, 30 novembre 2021.

LES CAHIERS DE L'ORDRE NATIONAL DES PHARMACIENS, « Numérique en santé, entre e-santé, exercices pharmaceutiques et usages, tout (sera) est connecté », n°18, juillet 2021.

BERNELIN, DESMOULIN & LEFEVRE (2020), « Données massives, big data et santé publique : de quoi parle-t-on ? », ADSP (Haut Conseil de la Santé Publique), n°112, septembre 2020, pp. 14-19.

ISAAC-SIBILLE (ASSEMBLÉE NATIONALE), « Rapport d'information déposé sur le dossier médical partagé et les données de santé », n°3231, 22 juillet 2020.

Quelques publications récentes sur l'IA dans le domaine de la santé :

https://www.actuia.com/actualite/lutilisation-de-lia-permet-daugmenter-le-nombre-de-depistages-et-de-suivis-de-la-retinopathie-diabetique/?mc_cid=889bef8ba2&mc_eid=29277ce73c ;

<https://www.nature.com/articles/s41746-024-01010-1> ;

<https://www.academie-medecine.fr/systemes-dia-generative-en-sante-enjeux-et-perspectives/> ;

<https://radiologybusiness.com/topics/artificial-intelligence/large-language-models-excel-simplifying-radiology-reports> ;

<https://www.who.int/fr/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models> ;

<https://observatoire-competences-industries.fr/etudes/edec-industrie-de-la-sante-3/> ;

<https://www.inserm.fr/dossier/intelligence-artificielle-et-sante/> ;

<https://op.europa.eu/en/publication-detail/-/publication/fb8d8ec2-55a0-11ed-92ed-01aa75ed71a1> ;

https://français.medscape.com/voirarticle/3609683?ecd=mkm_ret_231219_mscpmrk-ous_genmarketing-boy_etid6168532&uac=197766cy&impid=6168532 ;

<https://www.genengnews.com/topics/artificial-intelligence/machines-bring-efficiencyand-empathy-eric-topol-talks-ai-in-precision-medicine/> ;

<https://www.rtflash.fr/l-intelligence-artificielle-va-reinventer-medecine/article> ;

<https://medicalfuturist.com/7-things-to-expect-from-ai-in-healthcare-this-year/> ;

<https://www.lesechosleparisien-evenements.com/wp-content/uploads/2024/03/la-reco-2024-bd-bat-page-a-page.pdf>

L'évolution du commerce grâce à la numérisation

Par Leonardo ÁLVAREZ
Fenicio

Le commerce électronique représente un élément clé des processus de numérisation de notre société et un facteur majeur de croissance économique. Après avoir mis le sujet dans une perspective historique, nous définirons le concept général de commerce numérique et présenterons sa réalité d'aujourd'hui. Nous réfléchissons aux facteurs qui ont contribué à la croissance de cette industrie en analysant le contexte qui catapulte son développement actuel et futur. Enfin, nous réfléchissons aux tendances, solutions et concepts qui semblent dessiner l'avenir de cette industrie et nous ouvrons un débat concernant les acteurs qui éventuellement contrôleront le commerce numérique à l'avenir.

UN APERÇU RAPIDE DE L'HISTOIRE RÉCENTE DU COMMERCE

Bien que le commerce en général et le commerce international en particulier remontent aux origines de l'humanité¹, il suffit, pour les besoins de cet article, de commencer l'analyse à la fin du XIX^e siècle et d'en présenter uniquement les éléments clés.

La fin du XIX^e siècle et le début du XX^e siècle sont caractérisés par une augmentation de la production favorisée par l'émergence de la révolution industrielle et par les nouvelles formes de communication et de transport, ainsi que par le développement de nouveaux services des systèmes bancaires et financiers. Ainsi, le commerce international a été multiplié par 3 dans les 4 décennies précédant la Première Guerre mondiale. Le positionnement colonial de l'Europe représentait à l'époque un accélérateur du commerce international.

Après les ralentissements forts en lien avec les deux guerres mondiales ainsi qu'aux crises financières de l'époque et l'application qui a suivi de politiques protectionnistes, commence une étape de reconstruction et de coopération internationale caractérisée par des accords internationaux qui réactivent l'activité commerciale mondiale, parmi lesquels le Plan Marshall (entre les États-Unis et l'Europe), les accords de Bretton Woods (entre la Banque mondiale et le Fonds monétaire international) et l'accord général sur les tarifs douaniers et le commerce qui ont défini les bases fondatrices de l'Organisation mondiale du commerce, elle-même née en 1995 comme une extension du GATT. Entre 1950 et 2010, le PIB mondial augmente d'un facteur proche de 10 et le volume des exportations augmente d'un facteur 30.

Un élément clé de cette évolution, amorcée dans les années 1960, est la généralisation de l'usage des technologies numériques. Par ailleurs, le processus de décolonisation de

¹ Plus proche de nous, les traces les plus anciennes de la Route de la Soie remontent à plus de 4 000 ans. Il convient de noter que ces routes véhiculent bien plus que des produits (techniques, concepts, théories...), tout comme le font aujourd'hui les routes du commerce numérique.

plusieurs pays d’Afrique et d’Asie a modifié de nombreuses dynamiques commerciales en lien avec l’entrée dans les échanges internationaux de nouveaux pays et le fait que la scène commerciale mondiale a connu la transformation du transport maritime basée sur ce qu’on appelle la révolution des conteneurs.

Les années 1980 et 1990 sont caractérisées par un autre facteur d’accélération : les changements structurels dans la régulation commerciale. On voit bien l’effet de la libéralisation, de la réduction des tarifs douaniers et des accords de libre-échange ; mentionnons à titre d’exemple l’Accord de libre-échange nord-américain (ALENA) de 1994. Au cours de ces années, la Chine a commencé son développement commercial sur la base des réformes économiques menées depuis 1978, renforcé en 2001 avec l’entrée du géant asiatique à l’OMC.

Nous arrivons ainsi à la fin du XX^e siècle et au début du XXI^e siècle, avec l’arrivée de nouveaux acteurs (Google créé en 1999, par exemple), et l’explosion des processus de numérisation et de commerce électronique. À partir de ce phénomène, les canaux de génération d’affaires se multiplient et commence l’essor d’entreprises non seulement multinationales mais aussi mondiales.

En un siècle, nous sommes ainsi passés de marchés essentiellement locaux et nationaux, avec quelques routes internationales, à une industrie mondiale interconnectée et avec des marchés fortement dépendants les uns des autres. Aujourd’hui, nous nous dirigeons vers une mondialisation qui vraisemblablement entre dans une phase de décélération qui imposera de nouveaux changements dans le commerce international.

Nous ne pouvons évidemment pas ignorer, dans l’analyse qui suit, le rôle déterminant que joue et surtout celui que jouera dans l’évolution du commerce l’Intelligence artificielle (IA) et notamment penser à celui de l’IA générative.

COMMERCE NUMÉRIQUE, DÉFINITION ET VOLUME

Le commerce numérique dans sa définition la plus large et la plus simple fait référence à l’achat de biens et de services sur Internet, c’est pourquoi on l’appelle généralement « commerce en ligne ». Il existe de multiples façons pour une personne ou une entreprise de commercialiser un bien ou un service en ligne : magasins virtuels ou canaux de vente en ligne B2C/B2B, applications, plateformes de services en ligne de toutes sortes, places de marché, solutions d’économie collaborative, entre autres.

La Figure 1 (*cf.* page suivante) montre la croissance interannuelle des 10 dernières années, ainsi qu’une projection future, du volume du commerce numérique, un indicateur qui révèle la dimension du phénomène.

Le commerce électronique mondial a été multiplié par 5 au cours des 10 dernières années et devrait croître de près de 50 % de plus au cours des 3 prochaines années. Pour les raisons évoquées dans la section suivante, notre intuition est que les taux de croissance dans les années à venir seront encore plus élevés.

FACTEURS DE CROISSANCE DU COMMERCE NUMÉRIQUE

L’analyse qui suit s’appuie sur deux visions spécifiques, la première reflète les généralités visibles de l’industrie du commerce électronique et la seconde l’expérience professionnelle personnelle nourrie par l’interaction avec près de 600 entreprises de vente au détail qui ont mis en place leurs canaux de vente en ligne avec la plateforme de e-commerce dans le cloud en mode SaaS (*Software as a Service*), appelée Fenicio eCommerce.

Commençons par les facteurs généraux qui ont un impact positif sur la croissance du commerce numérique mondial.

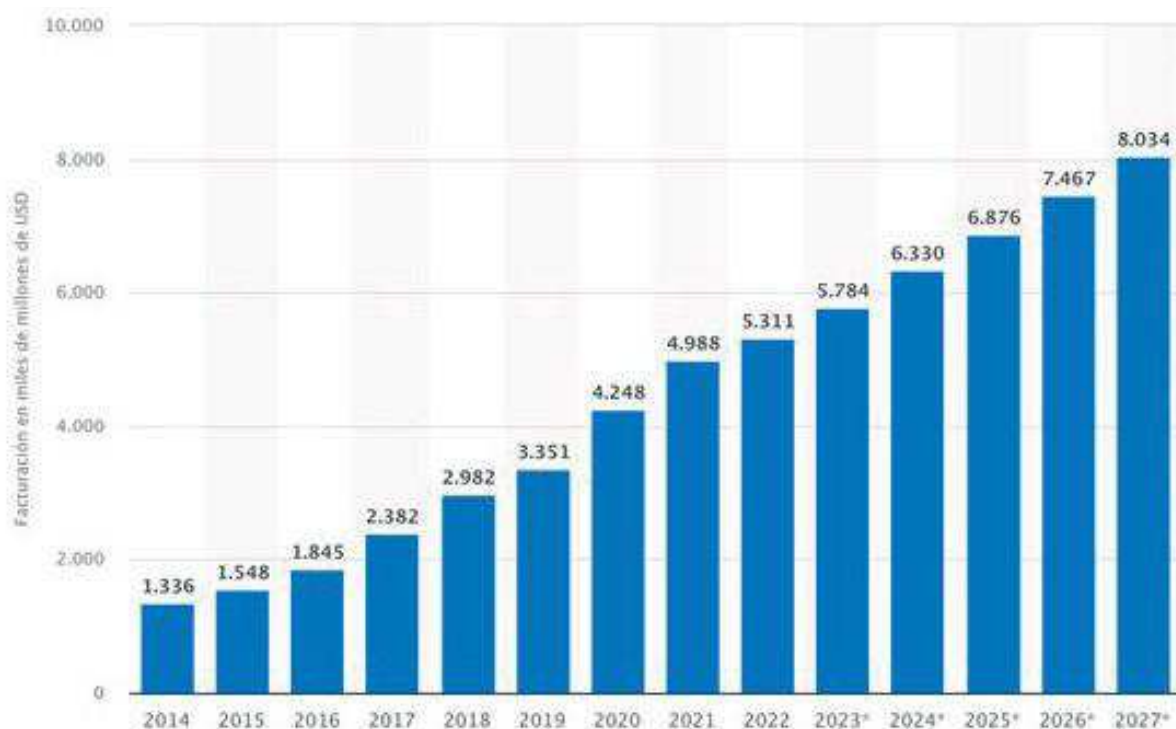


Figure 1 : Revenus des ventes par commerce électronique dans le monde entre 2014 et 2027, en milliards de dollars américains (Source : Statista).

Perméabilité technologique et infrastructures de communication

Les progrès des réseaux de communication dans le monde, récemment avec la 5G et les accès fibre, et l'émergence de projets d'accès massif à Internet par des réseaux non terrestres, tels que Starlink (SpaceX), Loon (Alphabet/Google) et OneWeb, entre autres initiatives, ont fortement démocratisé l'accès à Internet. Les investissements dans les infrastructures de réseau, la fibre optique sous-marine, les réseaux mobiles 5G et les multiples initiatives de soutien public ou public/privé de différents pays, ont amené Internet dans des zones géographiques reculées et donné l'accès à des centaines de millions de personnes à faible et même très faible revenu, en plus de générer une importante augmentation généralisée de l'utilisation de l'Internet.

Moyens de paiement

La modernisation et la simplicité des moyens de paiement ont accompagné et renforcé le développement du commerce numérique. Depuis les premiers pas où il était seulement possible d'effectuer des virements bancaires et du paiement à la livraison, en passant par la possibilité de paiement par carte de crédit et de débit, les grandes transformations ont été apportées par l'apparition des processeurs de paiement (passerelles qui facilitent énormément le développement de nouveaux sites de commerce électronique), des portefeuilles virtuels, des paiements mobiles, de la *Blockchain* et des crypto-monnaies. L'évolution a été constante en termes de sécurité, d'adaptabilité aux plateformes et de commodité pour le client final. En parallèle, les risques liés à la gestion de la monnaie virtuelle se sont multipliés, ce qui a également accéléré la recherche et le développement en matière de cybersécurité.

Services logistiques

Après l'effet multiplicateur de transactions généré par la pandémie, les prestataires logistiques du monde entier ont dû professionnaliser davantage leur offre et leurs processus pour pouvoir faire face à un nouvel ordre de grandeur de la demande et à une forte croissance des attentes en termes de qualité d'expérience. L'automatisation des intégrations avec les plateformes de vente en ligne pour fournir des niveaux de qualité adaptés aux nouvelles demandes des clients et l'intégration de services du dernier kilomètre qui résolvent les livraisons dans des délais minimum (1 à 2 heures) sont passées du statut d'options de services à haute valeur ajoutée à celui de service standard.

Autres technologies et services qui apportent de la valeur

Les solutions d'hyperpersonnalisation permettant de générer des expériences de navigation uniques pour chaque client ont commencé à jouer un rôle important dans les canaux de vente en ligne et ont élevé le niveau de concurrence. La nécessité d'optimiser les résultats commence à exiger une plus grande intelligence des campagnes de marketing par courrier électronique, ce qui rend nécessaire l'application de solutions d'automatisation axées sur des messages appropriés qui répondent aux attentes du client et améliorent considérablement le retour sur investissement de chaque contact.

Marchés mondiaux

Deux éléments sont ici importants :

- l'augmentation des achats en ligne transfrontaliers en raison de la croissance exponentielle des transactions en ligne de plateformes globales, telles qu'Amazon, Alibaba et plus récemment Temu, en plus de plusieurs marchés régionaux généralistes tels que Mercadolibre ;
- le développement massif des *marketplaces* verticales, dans de multiples secteurs d'activité.

Valorisation du confort en tant qu'attribut

Nous faisons surtout référence ici à la tendance forte à optimiser les temps de déplacement et à exploiter le bon usage du temps, caractéristiques très appréciées par les nouvelles générations de consommateurs.

Pandémie (Covid-19)

L'effet de la pandémie de 2020 a non seulement accéléré l'adoption du commerce numérique, mais a également amené tous les acteurs et participants de la chaîne d'approvisionnement à renforcer et à repenser leurs processus. Cette situation d'impact mondial nous a obligés à repenser bon nombre des pratiques déjà établies et à adapter les opérations à une croissance exponentielle de la demande.

Par la suite nous passons en revue les déclencheurs de croissance, tels que nous les comprenons à partir de la pratique et de l'expérience avec des équipes humaines dont la responsabilité est d'augmenter les ventes numériques et de maintenir des niveaux élevés de satisfaction des acheteurs en ligne.

Visibilité des résultats par la direction ou de la haute direction

Des résultats de vente spécifiques, avec des processus avancés d'analyse et des méthodes modernes présentation permettant de visualiser facilement des métriques multicritère, ajoutent de la confiance à la direction et ont généré des niveaux élevés d'implication de ceux qui ne croyaient pas à l'origine dans les canaux numériques pour leurs activités, mais qui voient les résultats des ventes augmenter d'année en année et surtout une explicabilité des résultats qui facilite la prise de décision stratégique.

Professionnalisation des ressources humaines

Les collaborateurs à l'origine des résultats ont été contraints de se former et de développer des compétences qui n'étaient pas nécessaires auparavant. Les rôles de responsables du commerce électronique, de *leaders* du Growth Hacking, de *leaders* de la conversion, de responsables de la réussite client et d'autres profils jamais requis auparavant, sont désormais nécessaires.

Nouvelles plateformes de commerce électronique en mode SaaS

Le logiciel en tant que service est arrivé dans différents secteurs et marchés pour y rester et le commerce numérique ne fait pas exception. La nécessité de rentabiliser les opérations commerciales et de professionnaliser les ventes en ligne à des coûts raisonnables a coïncidé avec la prolifération des plateformes de commerce électronique sous forme de logiciels en tant que service qui offrent une valeur élevée à des coûts directement liés au succès de l'entreprise. Jamais auparavant une technologie de e-commerce de classe mondiale n'avait été disponible à des coûts aussi abordables, démocratisant ainsi l'adoption de ces nouveaux canaux.

Engagement des gouvernements pour accélérer la numérisation

L'accent mis par les différents gouvernements, notamment ceux à économies moyennes ou émergentes, sur la promotion de la numérisation des entreprises (MPME et PME) a souvent représenté un déclencheur ou du moins un grand accélérateur. Des projets tels que celui de l'Agence nationale de développement de l'Uruguay (ANDE), appelé « Mode numérique », sont un exemple clair d'investissement de l'État dans le développement des capacités permettant une numérisation saine et progressive du secteur productif.

CE QUE NOUS ATTENDONS DU FUTUR PROCHE

L'utilisateur perçoit des changements qui semblent mineurs, comme le passage du PoS (*point of sales*), à la bande magnétique, puis à la puce, de la puce au sans contact, du sans contact au NFC et l'utilisation de *smartphones* et de solutions biométriques pour les paiements et donc, entre autres, le passage massif aux TPV (*third party verification*) avec de multiples options pour le payeur. Ce que l'on voit moins, c'est qu'au fur et à mesure de ces avancées, en parallèle, toute la chaîne d'approvisionnement du commerce numérique a travaillé et continue de travailler pour améliorer les technologies, les fonctionnalités et les expériences utilisateur, en mettant l'accent sur le client comme centre de toute l'action.

Même si l'histoire montre combien il est difficile de prédire les avancées dans le monde de la numérisation, certaines tendances semblent se consolider.

Avant de les présenter, mentionnons qu’au-delà de technologies et de solutions spécifiques, on voit émerger des changements de paradigmes à court et moyen terme. D’une part, nous passons du commerce électronique à tout un écosystème de transactions numériques qui va bien au-delà du commerce lui-même dans sa forme la plus directe d’achat/vente de produits ou de services (c’est-à-dire les transactions historiques de type produit ou service contre argent). Ceci est lié à ce que nous appelons aujourd’hui le *web 3.0* ou *web transactionnel*.

D’autre part, nous nous dirigeons vers une fusion de plus en plus forte entre commerce physique et commerce numérique, avec de nouvelles formes de présence physique, dont les magasins Amazon ne sont qu’un exemple. Ces solutions intègrent notamment des technologies de réalité virtuelle, augmentée et mixte. Par exemple, nos avatars nous permettront non seulement d’essayer virtuellement des vêtements, mais aussi de concevoir facilement les produits personnalisés que nous achetons. Nos créations, réalisées sans compétences particulières grâce à des solutions en ligne très ergonomiques, deviennent elles-mêmes des produits, que nous pouvons partager et éventuellement monétiser, et qui serviront de base à d’autres créations.

Nous indiquons maintenant quelques apports de la technologie à plus court terme.

Intelligence artificielle générative

La capacité de fournir des expériences uniques et personnalisées est devenu un objectif majeur dans le *e-commerce*. L’IA permettra de personnaliser l’expérience non seulement en analysant l’historique d’achat du client, mais également ses comportements et préférences individuels dans le monde en ligne et hors ligne. La disponibilité généralisée de services automatisés intelligents, *via* des *chatbots* et des moteurs conversationnels capables de dialoguer naturellement avec les humains grâce à l’application de l’IA générative, marquera une étape importante dans l’amélioration de l’expérience numérique du client.

Une autre contribution clé de l’IA générative sera donnée par la capacité du client à participer activement à la conception du produit qu’il souhaite acquérir.

Réalité augmentée, virtuelle et mixte

Bien que ces technologies aient déjà fait l’objet de tentatives d’exploitation dans le secteur du commerce numérique, nous n’avons pas encore vu d’exemples pratiques appliqués à grande échelle qui génèrent des résultats durables. La promesse de ces technologies est de fournir une expérience immersive imitant le monde réel ; cependant, leur application à grande échelle dépendra du fait que la production d’intrants soit une tâche raisonnable et réalisable. La possibilité pour le client d’essayer un produit dans le monde virtuel atteindra son expression maximale lorsque le client percevra réellement l’expérience de manière sensorielle et que son cerveau pourra être ainsi stimulé.

Cybersécurité et protection des données

Selon le rapport 2023 du Business Continuity Institute, qui définit les plus grands risques pour la continuité des activités des entreprises, la cybersécurité occupe la première place, suivie par les événements météorologiques violents et, plus loin, par les pannes des systèmes de télécommunications et les coupures des chaînes d’approvisionnement. Cette réalité implique un risque très important qui peut affecter la croissance du commerce numérique. Ces risques ne peuvent être combattus que par les entreprises, en définissant des politiques de développement sûres et en minimisant les risques en investissant sérieusement dans l’atténuation des problèmes de cybercriminalité, dont le coût est actuellement estimé à plus de 1,5 % du PIB mondial.

Blockchain et crypto-monnaies

La technologie *blockchain*, en tant que source d'authenticité dans la chaîne d'approvisionnement, peut agir comme un agent de réduction de la fraude. De plus, les paiements *via* les crypto-monnaies continueront d'augmenter et le système financier traditionnel devra y faire face et prendre une position plus claire face au carrefour qui se présente à lui. L'utilisation des *blockchains* permettra d'automatiser des transactions complexes, telles que l'achat et la vente de biens immobiliers, puisqu'une traçabilité numérique et sécurisée de tout l'historique notarié des produits sera possible, ainsi que des connexions instantanées avec les autorités juridiques et successorales correspondantes et avec des services tels que le cadastre.

Distribution disruptive

Le monde de la logistique dispose de multiples opportunités d'innovation et est ainsi confronté au défi de la transformation. À titre d'exemple, les drones et les véhicules autonomes proposeront de nouvelles façons de déplacer les produits, réduisant les délais de livraison et dans certains cas les coûts, mais ajoutant de nouvelles complexités liées aux infrastructures physiques nécessaires. Dans ce cadre, la mobilité électrique amène de nouvelles discussions auxquelles nous devons faire face, telles que la vitesse de recharge des véhicules et les formes d'élimination des batteries. Comme dernier point lié à la distribution, mentionnons l'avancée des centres logistiques automatisés, de plus en plus décentralisés, et de la robotique, apportant potentiellement une automatisation totale de l'ensemble du processus de distribution.

Connectivité mobile des objets

La communication mobile continuera d'être le dénominateur commun d'accès pour différents publics et pour les objets et machines, ce qui accélérera le monde des solutions de paiement mobile et des appareils ayant accès à Internet. Les objets du monde réel interagiront avec toutes sortes de systèmes pour s'adapter à nos besoins sans aucune action de notre part. Un exemple simple serait le véhicule électrique qui irait recharger sa batterie par sa propre décision et qui serait en charge du paiement correspondant. Les généralisations de ce concept à tous les secteurs d'activité sont innombrables.

Réduction de la chaîne d'approvisionnement

Dans un scénario futur possible, les grands fabricants établiront des liens commerciaux directs entre leurs usines et chacun des acheteurs potentiels, partout dans le monde. Cela constituera un défi majeur pour les intermédiaires qui devront continuer à innover pour diminuer l'impact du phénomène, au travers de services différenciateurs évolués. Les entreprises qui décideront de raccourcir leur chaîne d'approvisionnement par ce biais devront également innover pour réussir à tirer les bénéfices de ce raccourcissement de la chaîne d'approvisionnement.

Commerce unifié et commerce conversationnel

Bien qu'il s'agisse de deux concepts différents, le commerce unifié se concentre sur l'expérience client qui est l'axe de tout le processus d'achat, depuis la recherche du produit, la génération de la commande, la livraison, le paiement et le retour, dans le but de réaliser un parcours client extrêmement fluide, sans bruit ni problème. En termes de commerce conversationnel, les besoins d'attention 7 x 24 x 365 sans perte de qualité d'expérience deviendront la norme.

Durabilité et respect de l'environnement

Au fil du temps, les nouvelles générations évalueront de plus en plus la proposition de valeur en réfléchissant à l'empreinte environnementale et aux mesures palliatives que les entreprises vont mettre en œuvre. La sensibilité soulevée au problème climatique se retrouvera dans l'élaboration de l'avenir du commerce numérique.

Analyse de grands volumes de données

De nouveaux spécialistes dans le domaine seront nécessaires pour réaliser des exercices d'abstraction et appliquer le *big data* et l'IA dans le but de prendre des décisions basées sur l'analyse avancée de grands volumes de données.

Achat sur des plateformes multiples

Progressivement, tous les systèmes qui captent l'attention du public incluront une plateforme de *e-commerce* : les réseaux sociaux et les *metaverses*, les téléviseurs, les applications bancaires, etc. Ils vont générer de nouveaux modèles économiques qui compléteront leur activité et, dans certains cas, représenteront un élément critique pour le maintien sur le marché. Les téléviseurs deviendront de nouveaux espaces d'achat et de réapprovisionnement en produits pour la maison. Les super applications bancaires inciteront le client à gérer à travers elles toutes leurs opérations quotidiennes, y compris de vente et achat.

Intégration des canaux et concept d'omnicanal

Le commerce en ligne et hors ligne convergera de manière réelle et il n'y aura plus de dualité dans l'expérience d'achat du client, lui permettant d'acheter dans chaque cas sur le canal qui lui convient le mieux, sans perdre le contrôle et avec une fluidité des données, indépendamment du canal choisi.

EN CONCLUSION, LA DÉFENSE DU CONSOMMATEUR

En guise de synthèse, disons que le commerce numérique s'intégrera dans la société de telle manière qu'il deviendra transparent vis-à-vis des acteurs, en fournissant la plus grande efficacité pour la vie quotidienne. Son application mutera en fonction de l'évolution de ses divers composants : la technologie, l'expérience d'utilisation, les moyens de paiement et la chaîne logistique, entre autres.

La question que nous laisserons posée est de savoir si le client sera roi, notamment dans le cadre général d'une IA pas ou peu interprétable. Qui fixera les besoins et définira la feuille de route des innovations ? Les citoyens deviendront-ils simplement spectateurs d'une réalité qui sera définie et contextualisée par les prochaines intelligences artificielles à venir ? La feuille de route est ouverte, pour l'instant, c'est à nous tous, acteurs publics, privés et citoyens, de faire le nécessaire pour l'orienter de sorte à ce qu'elle soit compatible avec les modes de vie souhaités. Les réglementations nationales et internationales devraient jouer ici un rôle central, car la protection des consommateurs, si difficilement mise en œuvre, est confrontée à des défis beaucoup plus complexes dans cette nouvelle ère.

Le rôle de l'IA dans l'évolution des infrastructures numériques du futur

Par Francis JUTAND
Conseil général de l'Économie

Et Daniel KOFMAN
Télécom Paris

Face à la forte croissance de la complexité et à la dynamique, même à des échelles de temps très courtes, des infrastructures du numérique, l'intelligence numérique va apporter des outils critiques pour leur modélisation, leur conception, leur planification, leur surveillance, leur analyse, et leur exploitation, avec la création dynamique et à la demande de nouveaux services et applications, et donc pour le pilotage global du fonctionnement de ces infrastructures. La disponibilité de tels outils représente un “*game changer*” dans l'évolution des infrastructures et des services qu'ils rendent possibles. En ce sens, ils représentent aussi un élément clé de la souveraineté et un vecteur pour favoriser un positionnement fort de la France dans le contexte extrêmement concurrentiel que nous vivons. L'IA jouera donc un rôle clef dans la réglementation, l'innovation et la concurrence dans les infrastructures numériques, desquelles dépend plus que jamais le potentiel de développement de l'économie et de la société des pays.

L'INTELLIGENCE ARTIFICIELLE (IA)

Le terme Intelligence artificielle (IA) a été introduit dans les années 1950 pour définir un nouveau champ scientifique d'exploration des possibilités de réalisation technologique des capacités de l'intelligence humaine. Aujourd'hui, suite à certains succès marquants, il recouvre dans les médias et le vocabulaire commun un champ très large du numérique et notamment les robots et les automatismes informatiques, la reconnaissance et l'extraction de la parole et de l'image, le traitement automatique des langues naturelles, la traduction automatique, la modélisation de systèmes par apprentissage, les interfaces homme machine (IHM), ainsi que toutes les fonctions de génération de contenu (textes, voix, musique, images, vidéos, programmes informatiques), et cela notamment en mode interactif, comme pour les *chatbots* fournissant des réponses à des questions concernant de vastes champs de connaissances.

Si on s'intéresse à la partie de l'IA, qui lui a valu à l'origine le nom d'intelligence, des progrès ont été très rapides après 2010 avec l'utilisation de réseaux neuronaux, des grands modèles de langage (LLM pour *Large Language Models*) et les *transformers*.

Que ce soit au sens large de l'IA mentionné plus haut, ou plus restreint des IA fonctionnelles ou génératives et les traitements de données qui lui sont nécessaires, les IA sollicitent déjà massivement les infrastructures numériques. Ce phénomène s'accélère avec leurs nouveaux usages, mais aussi avec une distribution de plus en plus importante de l'IA, qui sort des grands centres de données pour se disséminer sur l'ensemble des

infrastructures, terminaux et objets du numérique. De nouveaux paradigmes, tels que les IA fédératives, ouvrent de nouvelles opportunités et accélèrent ce phénomène de diffusion

Ses usages devraient, comme le reste du numérique, progresser fortement à l'horizon des 10 ans.

Mais l'IA va jouer également un rôle important dans le développement et le pilotage des infrastructures numériques elles-mêmes, pour en maîtriser la complexité croissante et favoriser le plein usage de leurs potentiels, comme nous le verrons plus loin.

L'intelligence numérique

Concentrons-nous sur l'intelligence numérique, terme proposé pour parler du cœur de l'IA modélisatrice, fonctionnelle ou générative. Elle constitue le troisième étage de la fusée numérique après les communications ubiquitaires et le *cloud* et l'Internet des objets qui conjointement permettent d'observer le monde physique, de collecter ses données, de les stocker et de les traiter, ainsi que d'agir à distance sur ce même monde physique. Cette boucle peut s'appliquer à diverses échelles, du contrôle de la température de la maison aux grands écosystèmes planétaires en passant par les systèmes de production. Le troisième étage utilise les grandes masses de données, stockées ou collectées en temps réel, en les transformant en information, connaissances, contenus et décisions, et s'appuie sur les grandes capacités offertes par les deux premiers étages. Par ailleurs, ces divers étages fusionnent progressivement avec les avancées technologiques et les changements de positionnement des acteurs ; ce phénomène est accéléré par l'intelligence numérique qui développe les potentiels et accroît les capacités de l'ensemble.

À ses débuts, deux branches de l'IA sont apparues. D'une part, une approche basée sur la logique formelle, visant à décrire et formaliser les mécanismes de la pensée, d'autre part, l'approche neuronale, visant d'une certaine manière à émuler le comportement du cerveau. C'est la première qui l'a emporté dans les premières phases, mais c'est la deuxième, avec les réseaux de neurones artificiels et l'apprentissage profond, qui a permis les succès récents.

Dans un premier temps, on n'a pas pu atteindre les objectifs escomptés avec les réseaux de neurones qui requièrent l'accès à de très grandes masses de données et une puissance de calcul qui ne sont devenus disponibles que dans la dernière décennie, notamment grâce au *cloud* et aux composants adaptés aux algorithmes sous-jacents. D'ailleurs, ces nouvelles approches de calcul, bien au-delà des processeurs graphiques (GPU¹), sont en train de transformer profondément l'écosystème des producteurs de composants électroniques.

Les données existent depuis que l'on mesure ou utilise des capteurs et elles se sont numérisées. Les mathématiques appliquées, statistiques et probabilités ont développé des outils d'extraction de connaissance et de modélisation à partir des données. Avec la prolifération des données et les systèmes de *big data* permettant de les stocker et les traiter, on est passé de la logique et des mathématiques appliquées au *machine learning* ou apprentissage automatique, puis au *deep learning* avec la maîtrise de l'apprentissage de réseaux de neurones géants.

L'Intelligence numérique s'est ainsi développée à la rencontre de deux mouvements : la modélisation phénoménologique de systèmes par des réseaux de neurones et l'abondance de données. Elle est venue compléter la construction de modèles d'artefacts par analyse et assemblage de modèles de composantes et le développement du traitement de données par des approches statistiques et probabilistes.

¹ Citons par exemple les produits de l'entreprise Cerebras.

Finalement, dans divers domaines, il a été mis en évidence l'avantage d'une hybridation entre des modèles dynamiques décrivant le comportement d'un système (par exemple, sur la base d'équations différentielles, potentiellement stochastiques, ou des modèles décrivant des processus interagissant entre eux) et les approches statistiques et les modèles résultant des réseaux de neurones, souvent perçus comme une boîte noire, malgré les efforts en cours pour leur interprétation.

Les IA génératives

Les IA génératives représentent une innovation critique majeure pour faire face à la complexité grandissante en volume et diversité des données et connaissances. Du point de vue ontologique, les IA génératives donnent la possibilité de compléter les outils de raisonnement sémantique et causal opérant dans le champ temporel d'entrée sorties, par des outils statistiques.

Les IA génératives et leur développement massif reposent non pas sur une forme de modélisation première du comportement d'un système, mais sur les données et connaissances et informations diverses sur le système que l'on peut assembler et exploiter statistiquement de façon aveugle. Elles disposent ainsi de capacités d'exploiter des données non structurées, de faire ressortir des propriétés latentes, et de modéliser les réponses à des contextes et des questions posées par les opérateurs ou dispositifs automatiques de commande et de contrôle. Bien sûr la puissance de leurs usages réside dans le filtre des propriétés statistiques, mais il faut faire face en retour à des problèmes d'erreurs pour des réponses à des requêtes qui échapperaient au filet statistique et à sa modélisation par le *deep learning*.

On peut ainsi produire de nouveaux contenus, sur la base de contenus existants. À titre d'exemple, citons les systèmes de *generative design*, qui permettent de générer à partir de spécifications, une multitude d'objets y répondant. Rappelons par exemple que déjà en 2019 cette approche était utilisée en France pour concevoir des cadres de vélos. Plus récemment, les LLM capables de générer du texte, ont permis l'essor de solutions telles que ChatGPT, et les annonces d'IA capables de produire des images, de la musique, des vidéos et même des jeux vidéo fleurissent tous les jours.

Des outils pour aborder la complexité croissante des infrastructures numériques

Les infrastructures numériques sont en pleine évolution, comme cela est décrit dans diverses parties de ce document. La *softwarisation* des réseaux a permis une très grande flexibilité pour l'intégration de nouvelles fonctionnalités et pour la création dynamique, à la demande, non seulement de nouveaux services, mais également de réseaux grâce à la virtualisation qu'elle facilite amplement. L'orchestration efficace et efficiente de multiples fonctions, de multiples acteurs, dans le cadre des architectures convergentes réseau-*cloud*, pour la mise en œuvre dynamique de nouveaux services et applications, se traduit par une complexité accrue de la conception, de la planification, de l'exploitation et de la maintenance de ces systèmes, et l'IA peut apporter ici une contribution majeure. De même, l'intégration dans les infrastructures de nouveaux paradigmes et technologies, notamment radio, par exemple pour la mise en place des réseaux sans cellules (*cell-free*) ou de surfaces reconfigurables intelligentes pour une optimisation de l'usage du spectre, impose de nouvelles approches de conception et opération, hybrides entre modélisation mathématique et intelligence artificielle.

Sur ce potentiel de réalisation à la diversité colossale, il est nécessaire de construire des outils, des représentations, et de nouvelles modalités de conception et de contrôle pour faire face à la complexité grandissante.

Et tout cela dans un environnement contraint par la physique des communications, la sûreté de fonctionnement liée à la criticité des usages des infrastructures, ainsi que par des menaces de cybersécurité dont le champ potentiel croît avec la numérisation, mais aussi en prenant en compte la minimisation de l'impact environnemental.

Toujours face à une complexité qui croît exponentiellement, ces outils doivent être introduits dans un cadre qui facilite l'appréhension du contexte dans lequel se déroulent les événements à contrôler et qui simplifie les interactions homme-système. On voit ici l'intérêt des jumeaux numériques, comme cas d'usage particulier des concepts introduits dans notre autre article au début de la première partie de ce numéro. L'ensemble des choix à chaque instant étant impossible à évaluer par les méthodes actuelles, ces jumeaux numériques apportent la possibilité de simuler à large échelle l'impact de choix d'architecture ou des choix de configuration. L'IA joue un rôle clé dans l'orientation de ces simulations, afin de se focaliser sur des cas qui fournissent *a priori* le plus grand bénéfice potentiel. Ces jumeaux numériques utiliseront aussi d'autres types de modélisation, comme la modélisation stochastique et notamment la géométrie aléatoire, telles que présentées dans l'article de François Baccelli intitulé « Maîtrise de la complexité dans la 6G - L'approche par la géométrie stochastique » dans ce même numéro. L'hybridation mentionnée plus haut entre divers types de modélisation est ici essentielle et l'IA générative se présente comme un outil clé pour les interactions entre les opérateurs humains et ces jumeaux numériques.

LES USAGES DE L'INTELLIGENCE NUMÉRIQUE POUR LES INFRASTRUCTURES

On peut donc envisager, sous réserve d'une politique de qualification et de test approfondie et de modalités rodées d'interfaçages avec d'autres outils, la délégation aux IA d'opérations de conception, de planification, de surveillance, d'analyse, d'opération avec création dynamique de nouveaux services, et donc de pilotage global du fonctionnement des infrastructures numériques.

Il est évident qu'étant donné les standards de qualité hérités du monde des télécommunications, les "five nine"², et l'impact des pannes sur le fonctionnement des systèmes numériques critiques, récemment illustré, et des actions malveillantes comme les attaques de cybersécurité, la conception, le test et la mise en œuvre d'outils d'IA numériques vont être exploratoires, progressifs et donneront lieu à des politiques de sûreté et de résilience spécifiques, dont la validation reste aujourd'hui un problème largement ouvert.

Pour persévérer dans ce sens, il faut comprendre et apprécier les besoins pour des évaluations rapides de demandes de solutions utilisant l'adaptativité des infrastructures et leurs mises en œuvre opérationnelle, pour utiliser pleinement le potentiel d'ouverture des infrastructures (voir notamment la première partie de l'article), et proposer des réponses *ad hoc* de création de réseaux ou services dédiés, mobilisant des ressources tout en préservant des contraintes d'ensemble de réalisation et de sûreté de service.

Le potentiel des IA génératives peut être utilisé pour la création automatique de solutions et de services, ainsi que pour leur contractualisation : caractéristiques, performances, règles d'usage, tarification, engagements, pénalités, déni de service, résilience.

L'utilisation de l'intelligence numérique peut concerner la création dynamique de réseaux virtuels (potentiellement privés) et la mobilisation de ressources physiques de communication (notamment partagées), de stockage et de calcul, mais aussi les structures logiques de services, et les décisions d'intelligence des infrastructures pour implémenter par exemple

² Taux de disponibilité de 99,999 %.

des solutions de routage sémantique, orientés objectifs (voir l'article de Laurent Clavier et Marios Kountouris dans ce même numéro) et assurer la méta-résilience d'ensemble.

Plus prêt dans le temps, l'intelligence numérique jouera un rôle bien plus important qu'aujourd'hui dans la gestion d'automates pour les relations clients, la réalisation de transactions ou l'assistance personnelle. Mais aussi pour les terminaux eux-mêmes en les dotant de capacités d'interaction intelligente avec le pilotage des infrastructures physiques, de service ou d'intelligence.

LES PROBLÉMATIQUES D'USAGE DE L'INTELLIGENCE NUMÉRIQUE

Cependant, l'usage de ces technologies n'est pas sans poser de problèmes qu'il faut intégrer rapidement et précocement dans les attendus et les méthodes de développement. Ces problèmes sont de trois natures : la confiance, l'impact écologique et la souveraineté.

La confiance dans les usages

Le sujet de la confiance se pose de plusieurs façons : la confiance opérationnelle, la traçabilité et la perte de compétences.

Le sujet de la confiance opérationnelle a déjà été abordé plus haut et sera bien sûr au cœur des préoccupations dans le développement et la mise en œuvre de solutions d'intelligence numérique. Cet aspect est accentué par la multiplicité d'acteurs pouvant fournir simultanément des composants de solutions ou de solutions partielles à base d'IA et devant être orchestrés (voir la première partie de l'article).

La confiance de traçabilité concerne, d'une part, le fait que la modélisation par réseau de neurones n'offre pas de propriétés simples d'explicabilité et, d'autre part, vu l'impact majeur des jeux de données ayant servi à l'apprentissage sur les solutions proposées, les garanties pouvant être fournies sur les critères de choix de ces données et, plus amplement, sur les données effectivement utilisées, lesquelles ne peuvent souvent pas être fournies du fait de règles de confidentialité. Cet élément de la confiance se pose fortement dans un cadre d'insuffisance de solutions souveraines.

La confiance se construit en particulier par la capacité à valider les résultats de l'usage de l'IA. Dans ce cadre, des compétences avancées sont nécessaires, à la fois sur les systèmes traités (e.g. une installation industrielle) et sur les technologies d'IA utilisées. L'efficacité de l'IA peut, paradoxalement, engendrer une perte de compétences rendant impossible cette capacité de validation des usages de l'IA. Cette efficacité peut en effet amener progressivement à une perte de compréhension des systèmes, perte d'intérêt dans cette compréhension et une raréfaction des compétences en IA, par exemple, de statistiques et probabilités nécessaires pour pallier les problèmes et faire évoluer les modèles.

L'impact écologique de l'usage de l'intelligence numérique

L'impact écologique est à apprécier d'une part dans le cadre du développement des infrastructures numériques et de leurs usages et d'autre part dans les spécificités liées à l'usage de l'intelligence numérique.

Cet impact s'exprime en termes d'énergie consommée dans le fonctionnement des infrastructures numériques, leur fabrication et leur cycle de vie, ainsi que leur consommation d'eau et de ressources rares, entre autres. L'utilisation de l'IA du point de vue de la fabrication et de l'usage de ressources rares ne présente pas de spécificité qualitative par rapport au numérique en général. Par contre, l'IA est un outil important pour la

conception de solutions au problème mentionné, notamment dans sa contribution à la conception de nouveaux matériaux.

C'est du point de vue quantitatif, notamment avec l'avènement des IA génératives et leur usage, que se situe le problème de frugalité. Les propositions des GAFAM de développer les IA génératives sur des modèles géants de large usage ont un impact déraisonnable en termes de consommation d'énergie si on les utilise pour des applications spécifiques. Si l'on parle d'IA pour les réseaux, il n'est pas nécessaire de savoir traduire ou répondre à des questions de culture générale dans toutes les langues pour travailler à leur optimisation. Le développement d'IA spécialisées est la solution pour développer des usages frugaux dont les apports en économie d'énergie peuvent venir rendre du point de vue écologique, positif ou acceptable, l'utilisation de solutions d'intelligence numérique.

En matière écologique, si l'on se penche avec raison sur la consommation directe d'énergie apportée par le numérique en général et l'intelligence numérique en particulier, il faut aussi prendre en compte les usages par les entreprises pour la conception de produits et services qui apportent des gains en optimisation de consommation de matières ou d'énergie, ou l'utilisation de l'IA pour optimiser le fonctionnement de réseaux de toute nature ou de bâtiments : des gains significatifs et massifs mais aujourd'hui rarement mesurés et difficilement comptabilisables, tant ils sont dispersés.

La souveraineté des infrastructures

Le dernier point à aborder est celui de la souveraineté. Celle-ci est déjà plus que compromise au niveau européen en matière d'infrastructures de réseaux et de *cloud*, ce d'autant que les contraintes visant certains fournisseurs de technologies constituent une arme à double tranchant : en diminuant la concurrence elles peuvent ralentir l'innovation. La *softwarisation* des réseaux qui se poursuit ne fait que renforcer les problèmes de souveraineté en changeant la structure du marché (voir notre article intitulé « Des services réseaux aux plateformes du numérique et aux services multisectoriels » dans ce même numéro). Il en résulte dans ces domaines le besoin impérieux de renforcer les acteurs européens existants, favoriser l'émergence de nouveaux et agir pour la mise en place d'une puissance collective. Dans le domaine de l'IA et spécifiquement des IA génératives, il est donc vital de développer un écosystème européen d'offre et de demande souverain, sauf à perdre le contrôle des solutions de pilotage des infrastructures et à subir des failles de cybersécurité et d'intelligence dont l'impact socio-économique est incommensurable.

EN CONCLUSION :

L'INTELLIGENCE NUMÉRIQUE EST CLEF POUR LES INFRASTRUCTURES NUMÉRIQUES DU FUTUR

Du trafic et des fonctions nouvelles, les infrastructures numériques pour l'IA

L'IA et ses usages existent au travers des infrastructures numériques et imposent à celles-ci de nouvelles contraintes, certes en termes de capacité, mais aussi en termes d'urbanisation et de nouvelles propriétés fonctionnelles et non fonctionnelles. Tout d'abord le développement des IA élargit le champ des usages et accentue les besoins de capacité pour les échanges de données et de traitements massifs pour l'apprentissage et l'inférence. Mais les nouveaux usages demandent une diffusion accrue des capacités de calcul et d'intelligence, une nouvelle urbanisation des infrastructures numériques (voir la première partie de l'article) et apportent de nouvelles contraintes, notamment pour le contrôle de systèmes tels que ceux du transport ou des usines de demain, par exemple en termes de sécurité et de latence.

Leurs spécificités et les besoins de confiance dans leurs usages vont demander aux infrastructures la mise en place de nouvelles fonctions de traçabilité et d'audit pour les services développés à base d'intelligence numérique.

Des outils de rupture, l'IA pour la maîtrise de la complexité des infrastructures

Concernant le sujet du pilotage des infrastructures numériques du futur, l'intelligence numérique va apporter des outils critiques pour la planification, la surveillance, l'analyse, l'opération avec création dynamique de nouveaux services, et donc le pilotage global du fonctionnement de ces infrastructures. La disponibilité de tels outils représente un "*game changer*" dans l'évolution de long terme des infrastructures et des services qu'elles autorisent. En ce sens, ils représentent aussi un élément clé de la souveraineté et un vecteur pour favoriser un positionnement fort de la France dans le contexte extrêmement concurrentiel que nous vivons. L'IA jouera donc un rôle clef dans la réglementation, l'innovation et la concurrence dans les infrastructures numériques, desquelles dépend plus que jamais le potentiel de développement de l'économie et de la société des pays.

Du « sens » dans les réseaux

Par Laurent CLAVIER
IMT-NE

Et Marios KOUNTOURIS
EURECOM

Les réseaux de communications sont conçus et opérés sur la base de l'information telle que définie par Shannon, dépouillée de son sens et de son contexte. Si cette base a permis de faire évoluer les réseaux de façon incroyable, elle montre aujourd'hui certaines limites pour faire face à la fois aux applications toujours plus exigeantes et aux limites planétaires (à la durabilité). La théorie initiée par Shannon a permis le développement de solutions toujours plus performantes, mais sans regard sur la raison pour laquelle une information est transmise. La performance se mesure alors en débit, en efficacité – spectrale ou énergétique – mais pas en efficacité pour l'usage envisagé. Remettre le sens de l'information, la raison de sa transmission, au cœur de la conception et de l'orchestration des réseaux permettra d'atteindre les mêmes objectifs, ou même de remplir plus de fonctionnalités, tout en limitant, voire en réduisant, la quantité de données transmises et donc l'impact environnemental du réseau. *Less is more.*

HISTORIQUE ET MOTIVATION

Omniprésente, l'information se présente sous diverses formes, constituant un concept polymorphe et polysémantique. Claude Shannon, dans son ouvrage fondateur [1], a dépouillé l'information de son sens, de son contexte et de son interprétation. Ce faisant, il a rendu l'information mesurable. Il a délibérément considéré les aspects sémantiques comme sans rapport avec le problème de communication. Cette dichotomie entre information et signification est l'élément fondateur des systèmes de communication classiques ciblant une transmission de données à haut débit et sans erreur. Elle ne parvient plus cependant à répondre aux nouveaux besoins et aux contraintes de durabilité des réseaux émergents, 6G et au-delà.

Prenons l'exemple d'un très grand nombre de robots mobiles autonomes qui communiquent entre eux et cherchent un consensus visant à éviter les collisions. Atteindre cet objectif n'est pas une simple question de compromis débit-fiabilité-délai. La génération et l'échange de quantités excessives de données distribuées, en temps réel, parfois obsolètes, redondantes ou inutiles pour l'utilisateur final, vont saturer le réseau. Cette saturation risque d'empêcher le réseau de fournir le service attendu et soulèvera d'importants problèmes de sécurité pour les robots autonomes. Elle risque aussi de générer une augmentation exponentielle de la consommation d'énergie pour faire face aux volumes de données et aux contraintes de fiabilité anticipés. Pour une opération robuste et performante, il devient crucial de prendre en compte dans le processus de communication l'urgence et la valeur des messages fournis et de prioriser les transmissions pour satisfaire au mieux les exigences des applications.

Dans ce contexte, la Communication Sémantique Orientée vers un Objectif (CSOO), a récemment attiré une attention considérable. Elle introduit un nouveau paradigme de gestion de l'information qui permet une utilisation parcimonieuse des ressources de

communication et de calcul. Cette vision remonte à l'introduction du modèle de communication de C. Shannon écrite par W. Weaver [2]. Le problème de la communication est présenté sur trois niveaux :

- niveau A : avec quelle précision les symboles de communication peuvent-ils être transmis ? (le problème technique) ;
- niveau B : dans quelle mesure les symboles transmettent-ils précisément la signification souhaitée ? (le problème sémantique) ;
- niveau C : dans quelle mesure la signification reçue permet-elle la conduite souhaitée ? (le problème de l'efficacité).

Diverses tentatives ont visé la mise en œuvre du niveau B (communication sémantique). Les efforts les plus importants incluent les approches de logique probabiliste [3] [4], la théorie de la complexité [5], la théorie des jeux et le codage sémantique [6] [7]. Le niveau C a été abordé en utilisant les concepts d'information pragmatique [8] et de valeur de l'information [9] [10]. Néanmoins, le niveau A, le modèle de communication de Shannon, est, dans la pratique, resté incontesté. Aucune des extensions proposées n'a jamais été reconnue comme une théorie générale de l'information sémantique. Toutes les tentatives sont restées à un niveau conceptuel et n'ont pas abouti à une théorie élégante avec des applications immédiates, notamment dans les réseaux de communication.

Un nouvel élan pousse aujourd'hui la quête d'une théorie de la CSOO [11] [12]. Cet élan est motivé par les attentes, toujours plus grandes, placées dans les réseaux (débit, densité, fiabilité, latence) qui se confrontent aux exigences de durabilité. Qui plus est, cet élan peut aujourd'hui s'appuyer sur les développements envisagés pour les réseaux du futur. En particulier, les travaux autour de *communication and sensing* [13] [14] qui permettront une connaissance poussée de l'environnement en vue d'une meilleure prise en compte du contexte ; et la distribution de l'intelligence [15] [16] à tous les niveaux des réseaux, jusqu'aux objets faisant l'acquisition des données.

LA SÉMANTIQUE DE L'INFORMATION

Un élément indispensable pour libérer le potentiel de la CSOO est une définition concise, opérationnelle et universelle (si possible) de la sémantique de l'information (SoI). La SoI doit aller au-delà des mesures de substitution proposées jusqu'à maintenant telles que l'âge de l'information (AoI) [17] [18], l'âge de l'information incorrecte (AoII) [19], la qualité de l'information (QoI) [20] ou la valeur de l'information (VoI) [9] [10]. Ces mesures se limitent à certains attributs de la sémantique, adaptés dans certains contextes, mais loin d'être suffisants. La définition d'une SoI permettra d'adapter le processus de communication et les fonctionnalités clés associées (par exemple détection, apprentissage, traitement) aux besoins de l'utilisateur (ou application) final.

Contrairement à son usage courant en linguistique, en logique ou en informatique (*web* sémantique, bases de données, ontologies, etc.), la sémantique est ici employée avec son sens étymologique, celui de signification. La sémantique est une mesure de la signification, de l'importance et de l'utilité des messages par rapport à l'objectif de l'échange de données.

En premier lieu, la SoI est liée à l'importance relative de différents événements de même probabilité. Imaginons deux événements rares, se produisant chacun avec une probabilité très faible ($p \ll 1$), mais l'un comporte un risque majeur tandis que l'autre n'est qu'une particularité. Formellement, ils fournissent la même quantité d'informations ($\log p$), mais l'information véhiculée par le premier est évidemment d'une importance plus élevée. Cette disparité peut être intégrée à l'entropie source, ce qui entraîne une entropie dépendante du contexte et des mesures dérivées.

En deuxième lieu, la SoI peut être définie comme une fonction multivariée non linéaire d'attributs qualitatifs innés (objectifs) et contextuels (subjectifs). Les premiers sont inhérents à la génération d'information, quelle que soit son utilisation, comme la fraîcheur (AoI) et la précision. Les seconds dépendent de l'utilisation et du contexte, les plus pertinents étant la ponctualité, l'exactitude et l'exhaustivité. Il est important de souligner la dualité selon laquelle l'information a une valeur en soi, en plus de sa valeur « utilitaire » dépendante du contexte. Par exemple, la précision d'une mesure de capteur a une valeur intrinsèque liée à la qualité de sa représentation d'un phénomène, alors que cette même mesure a une valeur « utilitaire » différente selon son contexte d'utilisation, les mesures précédentes et les exigences de l'application (par exemple, s'il surveille la température dans une maison ou dans une centrale nucléaire).

CHANGER DE PARADIGME

Jusqu'à présent, la conception et l'évolution des systèmes de communication ont été principalement motivées par une approche maximaliste, qui fixe des objectifs audacieux mais souvent difficiles à atteindre. Dimensionnées pour les contextes les plus exigeants, les solutions vont bien au-delà des besoins la plupart du temps et dans beaucoup d'endroits, générant une inutile surconsommation. À l'opposé, la CSOO pourrait être considérée comme une approche de conception minimaliste (*less is more*), prônant un changement de paradigme de performances extrêmes vers des performances durables, où l'efficacité est garantie tout en améliorant considérablement l'utilisation des ressources (bande passante, énergie, calcul...).

La SoI ne fait pas de la communication une fin en soi mais un moyen d'atteindre des objectifs précis. La communication sémantique constitue alors un véritable changement de paradigme qui permet d'apporter où et quand cela est nécessaire les informations permettant la réalisation de l'objectif.

Grâce à l'intelligence distribuée dans le réseau, la communication sémantique permet de n'acquérir, traiter et transmettre que le contenu significatif et pertinent par rapport à l'objectif. Le nombre de symboles inutiles, traités et envoyés, est drastiquement réduit. Autrement dit, suivant le mantra selon lequel « tous les bits ne sont pas égaux », la communication sémantique pourrait renforcer « l'efficacité de l'information » des futurs systèmes de communication, au sens de maximiser le nombre de bits d'informations utiles extraits et fournis par ressource consommée. La SoI permet aussi d'exploiter la synergie entre le traitement des données, la transmission de l'information et la reconstruction du signal source. Cette approche disruptive permet de concevoir de nouvelles technologies de communication qui adaptent l'ensemble du cycle de vie de la donnée – depuis sa génération jusqu'à son utilisation, voire son obsolescence – à l'importance et à l'utilité de l'information.

Mettre en œuvre cette approche nécessite de remettre en question, voire radicalement changer, plusieurs principes de conception dominants. Il faut en particulier développer de nouvelles représentations des données, axées sur les objectifs, et introduire des mécanismes de priorisation des informations, qui permettront d'effectuer : un filtrage sémantique pour supprimer la redondance en générant et en envoyant uniquement des informations utiles et pertinentes ; un traitement sémantique, tel que l'extraction et l'étiquetage de caractéristiques, qui permet la reconstruction partielle ou approximative d'un processus ; et un contrôle sémantique pour orchestrer avec agilité la génération et la fusion d'informations multi-sources et de qualités multiples pour une utilisation efficace des ressources.

Les approches actuelles sont entièrement quantitatives et n'utilisent pas de contexte. Au contraire, Kountouris et Pappas [12] adoptent une approche qualitative pour différencier les informations et diffuser leur valeur dans les différentes couches protocolaires

des réseaux. Ce changement radical capitalise sur la capacité des dispositifs intelligents distribués à contrôler l'acquisition des données *via* un échantillonnage actif, dans lequel les échantillons sont prélevés en fonction de la variabilité de la source et du taux d'innovation du processus sous-jacent (la dynamique du phénomène à observer). De cette façon, seuls les échantillons « les plus informatifs » sont sélectionnés pour la transmission afin de répondre aux exigences sémantiques de l'application.

Cette approche structurellement nouvelle et synergique a le potentiel de découvrir des liens profonds et surprenants entre la théorie de l'échantillonnage, le traitement du signal parcimonieux, le calcul distribué, l'apprentissage et la théorie de la communication.

QUELQUES PISTES DE RECHERCHES

Baser la conception des réseaux sur une SoI pose en revanche de nombreuses questions de recherche. Nous en discutons certaines ci-dessous.

Métriques sémantiques

Un défi majeur consiste à établir des métriques et des mesures d'information sémantique concrètes, intégrant les attributs qualitatifs de l'information dans l'édifice théorique de la communication. Ces nouvelles mesures doivent capturer la dynamique des sources et du réseau, les interconnexions, non triviales, entre les attributs de l'information et les paramètres du réseau.

Accès multiple sémantique

Pour utiliser de manière optimale le medium partagé, les dispositifs doivent adapter leurs modèles d'accès non seulement au trafic exogène généré par les autres nœuds, mais également à la variabilité de la source ou du processus observé et donc à la sémantique de l'information, incluant les objectifs des applications.

Gestion des ressources axée sur les objectifs

Les informations multimodales multi-sources, souvent corrélées, acquises à différents niveaux de qualité sémantique, permettent d'atteindre les objectifs applicatifs de diverses manières. C'est une question à choix multiples de planification en temps réel, pour laquelle des algorithmes en ligne peuvent sélectionner quelle information, depuis où et quand, collecter et transmettre sous contraintes de ressources, de communication et de traitement.

Intégration de la communication et du contrôle

D'un point de vue de la théorie de l'information, il n'est pas clair de savoir comment les aspects de la communication orientée vers un objectif sont liés aux systèmes de contrôle en réseau avec des boucles de rétroaction sur les canaux de communication. Ces aspects peuvent être pertinents pour établir un cadre unifié intégrant les dimensions de génération, de communication et de contrôle des données.

Impact sur les protocoles et la couche OSI

Une mise en œuvre performante et efficace des systèmes de communication sémantique nécessite la génération de signalisations et de métadonnées spécifiques, reliant implicitement la couche physique à la couche applicative. Une conception et une utilisation efficaces des métadonnées sont clés pour obtenir des gains de performances élevés. De plus, certains processus réseau peuvent être transparents dans certaines couches ou

nécessiter de contourner certaines couches OSI pour un fonctionnement en temps réel ou pour éviter une congestion.

Sécurité et vie privée

La signification de la donnée est au cœur de l'ordonnement des communications sémantiques. Une première conséquence est l'augmentation potentielle de l'impact des attaques qui peuvent directement modifier la décision attendue. Une seconde conséquence est que la donnée transportée peut contenir une information plus importante et que la valeur sémantique des flux est partagée sur le réseau, ce qui peut rendre ces flux plus vulnérables à des vols d'information.

Autres possibilités

Au-delà du profond changement structurel nécessaire pour mettre la SoI au cœur de la vie de la donnée, l'évaluation même de cette sémantique repose sur les capacités des réseaux à s'adapter au contexte, nécessitant également de poursuivre certaines pistes de recherche actuelles. L'intelligence distribuée du cœur aux extrémités du réseau : chaque élément doit avoir des capacités de calcul et de prise de décision, en particulier les objets de périphérie qui doivent aussi être très peu consommateurs pour garantir une longue durée de vie sans maintenance (comme changer une batterie). La perception du contexte : au-delà des techniques de communication et de perception intégrées, chaque élément du réseau doit être en mesure d'évaluer son contexte de communication, aussi bien l'environnement physique que l'état des connaissances acquises par ses différents interlocuteurs. L'association de ces deux aspects pousse vers le *Joint Computation, Communication and Sensing*, un élément certainement clé des futurs réseaux et de la connexion des intelligences.

POUR CONCLURE

Mettre le sens au cœur de l'organisation des réseaux représente un profond changement et donc une indéniable difficulté. L'introduction de la SoI pourrait se faire dans le cadre de l'internet des objets. Les réseaux, généralement déployés pour un contexte précis, peuvent plus simplement introduire un contrôle basé sur la sémantique [21]. Pour une vision plus globale, la définition de l'Open-RAN (Radio Access Network) introduit une interface d'orchestration du réseau (RIC – RAN intelligent Controllers) [22] où peut se mettre en œuvre la SoI. La résolution des défis mentionnés, et de quelques autres, fournira la technologie fondamentale pour une pléthore de services socialement utiles et durables, comme le transport autonome, la robotique grand public, la surveillance environnementale ou la télésanté.

BIBLIOGRAPHIE

- [1] SHANNON C. E. (1948), "A mathematical theory of communication", *Bell Labs Technical Journal*, vol. 27, n°3, pp. 379-423.
- [2] SHANNON C. E. & WEAVER W. (1949), *The mathematical theory of communication*, University of Illinois Press, Urbana.
- [3] BAR-HILLEL Y. & CARNAP R. (1953), "Semantic information", *British Journal for the Philosophy of Science*, vol. 4, n°14, pp. 147-157.
- [4] FLORIDI L. (2005), "Semantic Conceptions of Information", in *The Stanford Encyclopedia of Philosophy*, Winter.

- [5] BAO J. *et al.* (2011), “Towards a theory of semantic communication”, in *IEEE Network, Science Workshop*.
- [6] GULER B., YENER A. & SWAMI A. (2018), “The semantic communication game”, *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, n°4, pp. 787-802.
- [7] WILLEMS F. M. & KALKER T. (2008), “Semantic coding: Partial transmission”, in *Proc. IEEE Intern. Symp. on Inf. Theory (ISIT)*, pp. 1617-1621.
- [8] SHARMA B. D., MITTER J. & MOHAN M. (1978), “On measures of ‘useful’ information”, *Information and Control*, vol. 39, n°3, pp. 323-336.
- [9] STRATONOVICH R. L. (1965), “On the value of information”, *Izv. USSR Acad.Sci. Tech. Cybern.*, n°5.
- [10] HOWARD R. A. (1966), “Information value theory”, *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, n°1, pp. 22-26.
- [11] POPOVSKI P. *et al.* (2020), “Semantic-effectiveness filtering and control for post-5G wireless connectivity”, *Journal of the Indian Institute of Science*, vol. 100, n°2, pp. 435-443.
- [12] KOUNTOURIS M. & PAPPAS N. (2021), “Semantics-empowered communication for networked intelligent systems”, *IEEE Communications Magazine*, vol. 59, n°6, pp. 96-102.
- [13] DONG F. *et al.* (2024), “Sensing as a Service in 6G perceptive mobile networks: Architecture, advances, and the road ahead”, *IEEE Network*, vol. 38, n°2, pp. 87-96, March.
- [14] ZHOU S. *et al.* (2024), “Task-oriented wireless communications for collaborative perception in intelligent unmanned systems”, in *IEEE Network*, doi: 10.1109/MNET.2024.3414144.
- [15] XU H., SENG K. P., ANG L. M. & SMITH J. (2024), “Decentralized and distributed learning for AIoT: A comprehensive review, emerging challenges and opportunities”, in *IEEE Access*, doi: 10.1109/access.2024.3422211.
- [16] GARAVAGNO A. M. *et al.* (2024), “An affordable hardware-aware neural architecture search for deploying convolutional neural networks on ultra-low-power computing platforms”, *IEEE Sensors Letters*, vol. 8, n°5, pp. 1-4, May.
- [17] KOSTA A. *et al.* (2017), “Age of information: A new concept, metric, and tool”, *Foundations and Trends in Networking*, vol. 12, n°3, pp. 162-259.
- [18] PAPPAS N. *et al.* (2023), “Age of information: Foundations and applications”, Cambridge University Press.
- [19] MAATOUK A. *et al.* (2022), “The age of incorrect information: An enabler of semantics-empowered communication”, *IEEE Transactions on Wireless Communications*, vol. 22, n°4, pp. 2621-2635.
- [20] BISDIKIAN C. *et al.* (2013), “On the quality and value of information in sensor networks”, *ACM Trans. Sen. Netw.*, vol. 9, n°4, pp. 1-26, doi: 10.1145/2489253.2489265, July.
- [21] ZHANG C. *et al.* (2022), “Goal-oriented communications for the IoT and application to data compression”, *IEEE Internet of Things Magazine*, vol. 5, n°4, pp. 58-63, Dec.
- [22] MARINOVA S. & LEON-GARCIA A. (2024), “Intelligent O-RAN beyond 5G: Architecture, use cases, challenges, and opportunities”, *IEEE Access*, vol. 12, pp. 27088-27114.

What digital infrastructure for the future?

- 05 **Introduction**
Francis JUTAND & Daniel KOFMAN

CROSS-SECTIONAL ELEMENTS ON CONVERGENCE

- 09 **From network services to digital platforms
and multi-sector services**
Francis JUTAND & Daniel KOFMAN

We are entering a new phase of the digital transformation. Continuing a transformation in silos, by sector of activity, we are moving to a phase which opens the way to multi-sector services, transversal to the various verticals. The enormous potential for value creation has been identified and will continue to develop given the potential for innovation which is thus opened; the risks for the actors and for the sovereignty of the countries are multiple.

This evolution is made possible by a series of new paradigms, including the softwarization and virtualization of infrastructures, the progressive convergence between networks and the cloud, the evolution of interfaces, digital twins and AI, as well as the opening of networks and systems to external actors. This last point transforms digital infrastructures into service platforms enabling the design and establishment of new services in real time, orchestrations components from different players, to dynamically respond to specific and evolving needs.

In this article we present an integrated vision of these various developments.

- 15 **The cloud in the digital industrial value chain**
Quentin ADAM

In the age of digital transformation, France and Western Europe are facing a major economic revolution. Digital technologies, and cloud computing in particular, are profoundly transforming the way we create and distribute wealth. This digital transformation represents an unprecedented opportunity, but also a major challenge for France, which has to juggle technological innovation with digital sovereignty. Unfortunately, this subject does not always receive the necessary attention, and the major impacts of this revolution are often downplayed.

- 25 **From manufacturing to infrastructure:
implications for regulation**
Pierre-Jean BENGHOZI

In the space of a few decades, digital technology has become an essential component of our societies. The regulation of digital infrastructures, such as telecommunications networks and data platforms, represents a major strategic challenge. They require massive investment and raise complex issues of sovereignty, equitable access and the distribution of value. Regulation must also adapt to rapid technological and market change, while managing the systemic risks associated with

cybersecurity and network reliability. This article examines these challenges and suggests ways forward for effective and fair regulation.

**31 Convergence of digital infrastructures:
an economic perspective**

Laurent BENZONI

Telecommunications operators (TELCOS) operate on a territorial basis, whereas the major content and application providers (GAFAM) operate on a global basis. The result for TELCOS is smaller size, higher debt and lower profitability. With convergence, the GAFAMs are moving down the digital infrastructure value chain to offer integrated services and capture greater value. They are getting closer to TELCOS but staying away from the local loop. Each TELCO subscriber requires a considerable investment compared with the services of a GAFAM.

While TELCOS have failed to move up the value chain, the emergence of slicing provides an opportunity to offer differentiated connectivity and to rebalance the conditions of access to their networks and subscribers. However, compliance with net neutrality and the use of artificial intelligence to implement these offers will limit TELCOS' room for manoeuvre. Convergence of digital infrastructures: an economic viewpoint.

40 Water in the clouds

Sylvain BOUVERET, Aurélie BUGEAU,
Anne-Cécile ORGERIE & Sophie QUINTON

In this article, we look at the environmental impact of digital infrastructures through the particular prism of water-related issues in the data centres (DCs) that make up clouds. We attempt to draw up an overview of the use of water in DCs, structured according to direct uses, for cooling and humidification, and indirect uses for power generation, production and the end-of-life of electronic equipment. This overview is illustrated with figures, mainly from the cloud service providers themselves, to give some orders of magnitude on water consumption and trends. The aim is also to discuss these figures, the relevance of the usual indicators and the associated potential impact transfers, and to recall the major spatio-temporal issues and conflicts of use linked to water resources.

48 A European digital perspective

Emmanuel DOTARO

Europe's digital future will be shaped by 6G, which, although it is a continuation of 5G, is characterised by the opportunity to bring together in an unprecedented combination disruptive innovations in all areas of digital technology. By understanding these components, technologies and architectures, we can identify the key issues that need to be addressed if Europe is to benefit from these transformations. The convergence between connectivity and computing, the uses of artificial intelligence, the integration of 3D architectures extending into space, cybersecurity and, lastly, the development of value through the control of digital consumption in the form of services are all areas in which Europe has a legitimate role to play and must be competitive. Economic development and the defence of societal values, where Europe is in the vanguard, can be achieved by combining investment in Research and Innovation, the federation of players and regulatory tools.

**TECHNOLOGIES AND ARCHITECTURES
FOR THE INFRASTRUCTURE OF THE FUTURE**

58 6G: promises and challenges for 2030

David GESBERT

This article takes a quick look at certain expectations of 6G that seem to have reached a consensus to date. We also shed light on the scientific concepts that will play an important role in these future wireless networks, such as intelligent reflective surfaces, artificial intelligence, spatialization of information and ISAC (integrated sensing and communication). Finally, we are addressing certain social challenges that will determine the success of 6G, such as computer security, the right to digital privacy, and the ecological impact of these wireless systems.

64 Communications networks on satellite constellations

François BACCELLI

This article, which is based on the Académie des Sciences report « Grandes constellations de satellites : enjeux et impacts », aims to describe the organisation, functionalities and prospects of communications networks using satellite constellations. It focuses on broadband communications functionalities. It describes the architectures used and the general characteristics of the satellites that make up these constellations. It also describes the principles governing communications between the various elements of the network. It reviews the longer-term prospects for this class of network.

**72 Managing complexity in 6G -
The stochastic geometry approach**

François BACCELLI

6G is based on the use of several new types of network elements, frequencies and architectures. Mastering the resulting complexity of radio coverage requires the development of new design and analysis tools. The aim of this article is to describe the principles of the stochastic geometry approach to these issues and the prospects for its use in the context of 6G.

**77 Cyber-physical systems and digital twins:
triggers and conditions for accelerated adoption**

Pascal BROSSET

The cyber-physical systems (CPS) terminology, which appeared contemporaneously with Industry 4.0, has been the topic of many academic publications about its potential in domains ranging from industry to healthcare. As a reminder, the term was introduced in 2006 by Dr Helen Gil of the US National Science Foundation. This concept was developed to describe the integration of computing with physical processes, where embedded systems and networks allow to monitor and control physical processes, including feedback loops between both.

The convergence of socio-economic factors and accelerating progress as far as artificial intelligence result in an acceleration in the adoption of CPSes, often bundled with digital twins. This article explores, from a manufacturer's point of view, the reasons of this acceleration and its implications.

84 The benefits of Gaia-X
Anne-Sophie TAILLANDIER & Pierre GRONLIER

Gaia-X, created in 2020 by the German and French governments, aims to create a federated data infrastructure guaranteeing data sovereignty and transparency, thereby supporting Europe's digital economy. With over 300 members, Gaia-X promotes interoperability and security through a standardised framework, facilitating data sharing while maintaining control and compliance with European regulations such as the RGPD. It emphasises technical, operational and legal autonomy, encouraging multi-cloud strategies to reduce dependency on non-European suppliers. The Gaia-X architecture relies on standardised vocabularies, encrypted verifiable references and a trust registry to ensure secure and verifiable data exchanges. The initiative also supports the development of AI by ensuring data traceability and consent, in line with regulations such as the AI Act.

Going forward, Gaia-X aims to integrate future regulations, providing comprehensive coverage for data providers and consumers, fostering a trustworthy digital ecosystem.

94 Virtualised networks: promises and challenges
Fabrice GUILLEMIN

Virtualisation technologies have revolutionised telecommunications networks in recent years. Aside from the technical advances and the considerable flexibility they bring, this breakthrough in network design poses a number of challenges, particularly with regard to the underlying cloud infrastructures and network operations. The aim of this article is to take stock of these sensitive issues for network operators and identify possible solutions.

99 The softwarization of networks and its technical and economic impact
Emmanuel PUIG

Softwarization, or the migration of IT functions from hardware to software, marks a key technological development in which increasingly powerful CPUs enable software solutions that were previously impossible. This transition, visible in a variety of fields such as video games and telecoms, is driven by economic and operational benefits such as flexibility, cost reduction, and access to a wide range of open source tools and resources. In telecoms networks, although network cores are already largely "softwarised", radio access is still predominantly hardware-based, but the trend is changing with the rise of private networks and technologies such as 5G, favouring more flexible solutions tailored to small volumes. Companies that embrace this transformation at the right time can gain a significant competitive advantage.

104 New optical technologies: what impact on tomorrow's augmented reality systems?
Jean-Louis de BOUGRENET DE LA TOCNAYE

Augmented reality (AR) has often progressed thanks to technological leaps. This has been the case with the miniaturisation of displays, pico-projectors and combiners. Will the latest advances in the design/manufacture of surface-effect optics (e.g. meta-surface, -lens and -shape) have the same impact on the projection devices of future AR systems? Progress has already been made in this area with freeforms. Similarly, will the development of holographic AR, driven by the growing

popularity of smart glasses in the consumer sector, impose this choice on all AR devices? These are all questions that will be briefly addressed in this article, which focuses on the optical layer of AR systems. Finally, in this ongoing race towards ever greater immersivity and closer proximity to the eye, will contact lenses be the future of hyper-immersive AR? What are the ultimate dimensions that can reasonably be crossed? These are just some of the questions we'll be tackling here.

113 **Metavers: beyond virtual reality headsets**

Guillaume MOREAU

Since Mark Zuckerberg's speech on the "metaverse" in 2021, it has to be acknowledged that the claimed generalization has not taken place yet. While immersive technologies have notably improved and reached price levels compatible with a wide audience, some challenges, more or less related to the tradeoffs on the interfaces, remain to be addressed. Not only must the software part adapt to the tradeoff between mass, energy, network bandwidth and computing power but authoring tools must also improve. Those systems are far from reaching human perception features and generate aftereffects that may have some consequences on human health. At last, it is important to study the ethical questions raising from those technologies that are mixing together with artificial intelligence.

119 **Cyber security**

Hervé DEBAR

Cybersecurity has become a major challenge for digital infrastructures. The threat carried out by attackers that are highly skilled, motivated and tooled leaves no doubt to the fact that digital infrastructures, current or future, will be under attack. Despite the best care, some of them will be compromised. Therefore, cybersecurity best practices need to be deployed in all these digital infrastructures, to understand and analyze risk, to protect these infrastructures from malicious activities, and to detect and mitigate the impact of these activities as early and efficiently as possible. While there has been progress in both cybersecurity methods and tools, we need to ensure that cybersecurity aspects are taken into account in the management and operation of these digital infrastructures. Ensuring this will increase the efficiency and trustworthiness of these digital infrastructures, and benefit all users and stakeholders.

127 **Components for digital infrastructures**

Dimitri KTÉNAS & Dr Denis DUTOIT

In this paper, we take a closer look at the components that make up digital infrastructures, first from the point of view of computational infrastructures, then from that of communications infrastructures.

Key components of the computational infrastructure of digital systems include intelligent sensors, application processors, and any xPU providing computational capacity. These components create a heterogeneous, flexible and scalable digital processing infrastructure, capable of meeting growing computing needs.

On the other hand, the telecommunications sector has become the keystone of the Western economy. It appears that 6G, currently being defined, will require new dedicated components, notably due to the rise in frequency, which is likely to render current silicon components inoperable. A real breakthrough is therefore on the horizon in this field of microelectronics.

SECTORAL IMPACTS

134 Contribution of the Agdatahub data sharing infrastructure to the agri-agro sector
Sébastien PICARDAT

France's farms are highly connected and generate a huge amount of data. The Agdatahub data-sharing infrastructure is the trusted third party for the agricultural and agri-food sector, enabling data to circulate in compliance with the European regulatory framework. The French platform is a technological and operational tool that meets the objectives of sustainability, innovation and sovereignty pursued by players in the sector. As a data intermediary, a status enshrined in European regulations and French law, Agdatahub simplifies and secures the circulation of agricultural data, and addresses priority use cases such as animal genetics, carbon reduction, environmental labelling, etc.

142 Digital challenges in electricity distribution
Claire WAAST-RICHARD, Claudine RABILLARD & Yves BARLIER

The power distribution network is, all over the world, at the crossroads of many transitions. The energy transition is already a reality and has a strong impact on the network which, in its turn, adapts. The digital is one of the key success factors. Deployments of smart meters, IoT, data processing, AI and simulation allow Enedis to generate performance; to support its customers in their energy transition; and allow the electricity system and market players to operate more precisely. The continued growth of decentralized renewable energy installations, the generalization of electric mobility, the development of hydrogen will have an increasingly profound impact on the way the network is managed. The future system will be much more complex and more decentralized. This will not happen in a day. The transition will be achieved thanks to digital technologies but also thanks to commitment of all the players in the sector, with a "smart regulation" and with all employees mobilization.

149 Cooperative intelligent transport systems for urban mobility policies
Éric MONCEYRON

Connectivity now makes it possible to integrate the vehicle with its immediate environment, and to provide new safety aids for the driver, including the promotion of peaceful cohabitation with other modes of transport, as well as new opportunities for road and traffic managers. This assistance is implemented in particular by Cooperative Intelligent Transport Systems (STI-C), based on the instantaneous exchange of data between vehicles (V2V) and infrastructures (V2I).

After intensive standardisation work, the European Commission launched a decade of successive pilot projects to deploy ITS-C. These projects have enabled us to come to terms with operational realities and have now reached the production stage, as demonstrated by the final seminar of the InDiD project (Digital Infrastructure of Tomorrow - June 2024).

This article highlights the contribution of ITS-C to the urban nodes of the trans-European transport network, integrating vehicles into reactive multimodal management of global mobility, particularly on the outskirts of city centres where the modal share of the car remains dominant.

**161 Pleading for the great forgotten:
public data-sharing infrastructures**
Laura LÉTOURNEAU

What if the infrastructures of the future were not so much those with spectacular underlying technologies as those that help us to meet the challenges of the future? Data-sharing infrastructures organise the fluid and secure circulation of information within a given circle of players. Ecological transition, pandemic management, improving the education system, security... Infrastructure of this kind is actually necessary whenever a large number of public and private players need to work together to solve societal problems. What could be more important? Yet the subject suffers from a deleterious asymmetry of attention in favour of “sexier” technologies such as AI. This paper explains why these infrastructures are essential and sets out a collective and ambitious guide to help support them, in the belief that we can collectively move faster in the next 3 years than we have in the last 20.

173 Healthcare systems
Michel BARTH

Digital technology is investing in and transforming the healthcare sector, and this trend shows no signs of slowing down, particularly with the development of high-speed internet, cloud computing, mobility and connected objects (Internet of Things or IoT). Beyond the obvious rationalization and efficiency of care enabled by the digitization of organizations and processes, digital health also brings a qualitative promise of increasing patients’ chances of being accurately diagnosed as early as possible, and then being treated and rehabilitated effectively, while streamlining care episodes, especially for hospitalizations and long-term or chronic conditions. All generated data is transmitted over fixed and mobile networks for various purposes in support of the patient care journey. This results in a continuously growing need for fixed and mobile digital infrastructures to support these flows.

184 The evolution of commerce thanks to digitisation
Leonardo ÁLVAREZ

Electronic commerce is a key element in the digitisation of our society and a major factor in economic growth. After putting the subject into historical perspective, we will define the general concept of digital commerce and present its current reality. We will consider the factors that have contributed to the growth of this industry, analysing the context that is catapulting its current and future development. Finally, we will reflect on the trends, solutions and concepts that seem to be shaping the future of this industry, and we will open a debate on the players who will eventually control digital commerce in the future.

FORWARD-LOOKING STATEMENTS

**192 The role of AI in the evolution
of the digital infrastructures of the future**
Francis JUTAND & Daniel KOFMAN

Faced with the rapid growth of complexity and the dynamic, even at very short time scales, of digital infrastructures, digital intelligence will bring critical tools for their modelling, design, planning, monitoring, analysis and exploitation, with

the dynamic creation and demand of new services and applications, and therefore for the overall management of the operation of these infrastructures. The availability of such tools represents a “game changer” in the evolution of infrastructures and services they make possible. In this sense, they also represent a key element of sovereignty and a vector to promote a strong positioning of France in the extremely competitive context we live. AI will therefore play a key role in regulation, innovation and competition in digital infrastructures, the potential for development of national economies and societies depends more than ever on these.

199 Meaning in Networks

Laurent CLAVIER & Marios KOUNTOURIS

Communications networks are designed and operated on the basis of information as defined by Shannon, stripped of its meaning and context. While this basis has enabled networks to evolve incredibly, it is now showing certain limitations in terms of coping with both increasingly demanding applications and planetary limits (sustainability). The theory initiated by Shannon has enabled the development of ever more efficient solutions, but without regard for the reason why information is transmitted. Performance is measured in terms of throughput and spectral or energy efficiency, but not in terms of effectiveness for the intended purpose. Putting the meaning of the information, the reason for its transmission, at the heart of network design and orchestration, will make it possible to achieve the same objectives, or even to fulfill more functions, while limiting, or even reducing, the quantity of data transmitted and therefore the environmental impact of the network. Less is more.

Issue editors

Francis JUTAND and Daniel KOFMAN

Ont contribué à ce numéro

Quentin ADAM est un entrepreneur du numérique. Il est le CEO de Clever Cloud, une entreprise qu'il a fondée en 2010, et qui s'est imposée comme une référence incontournable dans le domaine du *cloud* en Europe. Clever Cloud, en forte croissance ces dernières années, se distingue par ses innovations dans les technologies PaaS et sa capacité à fournir des solutions robustes et efficaces pour augmenter la productivité des développeurs.

Il est très impliqué et reconnu pour son engagement en faveur de la souveraineté numérique européenne et de la communauté *open source*, et est le président de l'Open Internet Project (OIP), association rassemblant des entreprises qui s'engagent activement à protéger les intérêts numériques de l'Europe et à renforcer sa position sur la scène mondiale. L'association OIP se donne pour mission de promouvoir un internet ouvert et équitable, en soutenant des politiques qui favorisent l'innovation et la concurrence dans le secteur numérique européen.

Quentin Adam est également co-fondateur du respirateur artificiel *open source* MakAir, un projet novateur lancé en réponse à la pandémie de Covid-19 qui a permis de fédérer, de mobiliser des ressources diversifiées et d'innover dans des secteurs critiques de la santé. Conférencier actif, il défend une vision de l'informatique, des données, de l'économie, tant d'un point de vue technique que stratégique, et promeut une vision d'entrepreneur engagé dans une vision d'une France et d'une Europe souveraine en créant un environnement numérique ouvert, durable et collaboratif.

→ *Le cloud dans la chaîne de valeur industrielle du numérique*

Leonardo ÁLVAREZ est entrepreneur, passionné par le changement et par la recherche constante d'opportunités permettant de développer des solutions et des modèles économiques sur Internet. Depuis 2002, il est partenaire d'Análoga, actuellement Fenicio eCommerce (www.fenicio.io), une plateforme de commerce électronique dans le *cloud*, *leader* sur le marché uruguayen et axée sur les entreprises B2C pour le segment de vente au détail de plusieurs catégories d'articles (mode, électronique, maison, meubles, alimentation, entre autres).

Actuellement, sa tâche principale est de développer l'expansion commerciale de Fenicio eCommerce en Amérique latine et dans les Caraïbes, en multipliant la présence de Fenicio au Paraguay, au Pérou, au Mexique, au Chili, en Argentine et en République dominicaine. Aujourd'hui, il travaille à l'expansion de son activité au Paraguay et au Pérou.

→ *L'évolution du commerce grâce à la numérisation*

François BACCELLI est directeur de recherche à l'Inria et professeur invité à Télécom Paris. Il est actuellement responsable de l'ERC advanced NEMO sur la dynamique des grands réseaux. Il a été titulaire de la chaire Simons Math+X à UT Austin, États-Unis, entre 2011 et 2020. Il est membre et co-fondateur du LINCS.

Ses recherches se situent à l'interface entre les mathématiques et les réseaux de communications. Ses travaux mathématiques se concentrent sur les processus ponctuels, les graphes aléatoires et la géométrie stochastique d'une part et la dynamique des réseaux d'autre part, notamment la dynamique des systèmes de particules et des réseaux de files d'attente. Ses principales contributions dans le domaine des communications sont centrées sur la théorie de l'information, le contrôle de congestion et la modélisation des réseaux de communications cellulaires.

François Baccelli est *docteur honoris causa* de l'Université Heriot-Watt d'Édimbourg, Royaume-Uni (2016). Il a reçu le Sigmetrics Achievement Award de l'Association for Computer Machinery aux États-Unis en 2014, les prix Stephen O. Rice et Leonard G. Abraham de l'IEEE Communications Society en 2014, ainsi que le Grand prix France

Télécom 2002 de l'Académie des sciences. Il est membre de l'Académie des sciences depuis 2005.

→ ***Réseaux de communications sur constellations de satellites***

→ ***Maîtrise de la complexité dans la 6G - L'approche par la géométrie stochastique***

Yves BARRIER a plus de 30 ans d'expérience dans le secteur électrique, aussi bien dans ses dimensions techniques qu'économiques, tarifaires et juridiques. Actuellement directeur planification des réseaux et *smart grids* d'Enedis, le plus gros distributeur d'électricité européen, il était auparavant responsable des relations avec les services ministériels et la Commission de Régulation de l'Énergie pour Enedis.

Il a également travaillé en ingénierie et en exploitation des réseaux électriques, sur l'optimisation de l'équilibre offre demande et sur l'insertion des énergies renouvelables dans les systèmes électriques. Il est ingénieur, diplômé de Supélec en 1987.

→ ***Les enjeux du numérique dans la distribution d'électricité***

Michel BARTH, Polytechnicien (1979), diplômé de Télécom Paris (1984), est président et co-fondateur d'ENoving. Il a une expérience approfondie du conseil en stratégie acquise notamment au sein d'un grand cabinet international avant de fonder ENoving. Il a travaillé sur de grands projets de stratégie et conduite du changement dans des secteurs d'activité variés avant de se tourner vers la santé et le médico-social, où il accompagne les entreprises et organisations publiques dans leurs réflexions stratégiques et leurs programmes de croissance et d'innovation.

Il a par exemple aidé l'Institut Mines Télécom dans le développement de son plan de recherche et innovation en santé, ayant conduit à des partenariats avec de grandes entreprises et au montage de chaires comme celle sur l'innovation numérique dans le bloc opératoire avec l'AP-HP. Il a aidé des *start-ups* et divers organismes dans leurs projets d'innovation en santé. Il a accompagné sur plusieurs années des régions françaises dans le montage et la mise en œuvre de plateformes de télémédecine (téléconsultation, télé-expertise et télésurveillance médicale) ainsi que pour l'échange et le partage d'images. Il a également participé à plusieurs dépôts de demandes de subvention notamment pour l'AMI CMA ayant tous conduit à des succès et des projets en cours de déploiement à ce jour.

Par ailleurs, Michel Barth est formateur sur des sujets liés à l'entrepreneuriat, au *business modeling*, au *design thinking* ainsi qu'à la santé numérique. Il est ainsi chargé de cours à l'École Centrale (Paris & Pékin), supervise des thèses professionnelles dans le cadre de Mastères à HEC Executive Education et Dauphine (master santé), et intervient à ESME, UTC, ESIEE, ENSIEE ainsi qu'à l'EHESP. Il est expert auprès d'EIT Health en Europe. Il coanime un *think tank* d'anciens élèves de l'X sur l'innovation en santé (X Santé). Il a également participé et réalisé de nombreuses publications dans le domaine des modèles économiques et de l'innovation.

→ ***Les systèmes de santé***

Pierre-Jean BENGHOZI, directeur de recherche émérite au CNRS et à l'École polytechnique, a une triple formation d'ingénieur comme ancien élève (X76), de docteur en Sciences des organisations et d'HDR en économie de l'Université Paris-Dauphine. C'est un des précurseurs, reconnu internationalement, des recherches sur les modèles économiques du numérique dans les industries culturelles et créatives. Il a dirigé, jusqu'en 2013, le pôle de recherche en économie et gestion (UMR X-CNRS) et la chaire « Innovation et régulation des services numériques » qu'il avait fondée en partenariat avec Telecom ParisTech et Orange. Il a été ensuite directement impliqué dans la régulation du secteur comme membre du collège de l'Arcep, membre du comité de prospective de la Cnil et comme président de la mission d'évaluation du Plan France Très Haut Débit réalisée à la demande du Gouvernement et de la Commission européenne. Il est régulièrement

sollicité sur ces questions, au niveau national et international, auprès d'institutions publiques comme d'entreprises privées.

Pierre-Jean Benghozi a publié plus de 200 contributions académiques et enseigne régulièrement dans plusieurs grandes universités parisiennes et étrangères, notamment comme professeur à l'Université de Genève de 2019 à 2021 ou au Business Science Institute (Luxembourg) depuis 2014. Comme enseignant-chercheur, il s'est aussi profondément impliqué dans l'administration de la recherche : direction de laboratoire, présidence de section du Comité national de la Recherche Scientifique, présidence de la Société Française de Management, conseils scientifiques...

→ ***De la manufacture à l'infrastructure : des conséquences pour la régulation***

Laurent BENZONI est Docteur ès sciences économiques et professeur agrégé des facultés de Droit et de Sciences économiques. Il a créé et dirigé le département de Sciences économiques et sociales de Telecom Paris de 1986 à 1995. Il rejoint ensuite l'Université Paris-Panthéon-Assas où il est en poste, rattaché au Centre de Recherche en Économie et Droit.

Il a créé en 1995 TERA Consultants, un cabinet d'expertise économique et de conseil intervenant dans les domaines de la régulation, de la concurrence, de l'évaluation des dommages économiques et des études d'impact. Les méthodes originales, la pluridisciplinarité alliant compétences techniques, économiques et comptables valent à TERA Consultants, basé à Paris, d'avoir déjà réalisé plus de 1 500 missions dans 56 pays sur 5 continents.

→ ***Convergence des infrastructures numériques : un point de vue économique***

Jean-Louis de BOUGRENET DE LA TOCNAÏE est Professeur, chef du département d'Optique à IMT Atlantique, spécialiste des cristaux liquides composites, de l'optique diffractive et de la vision. Alexander von Humboldt Fellow, Churchill College Fellow (Cambridge) et Optica Fellow. Auteur de plus de 65 brevets, de 180 articles de rang A, fondateur de six entreprises. Lauréat du Cambridge Entrepreneurship (2002), du SPIE Technical Achievement Award (2006) et du Grand Prix de l'Académie des Sciences (2022). Actuellement directeur scientifique de la société Cylensee basée à Pittsburgh (Pennsylvanie) au sein du Vision Institute de l'UMPC dirigé par J.A Sahel, il conçoit et élabore des lentilles de contact instrumentées pour divers domaines de l'optique et de la santé. En particulier, il s'intéresse à la prévention de la myopie précoce, et à la réhabilitation sensorielle visuelle, liée à différentes formes de dégénérescences maculaires.

→ ***Nouvelles technologies optiques, quel impact sur les systèmes de réalité augmentée de demain ?***

Sylvain BOUVERET est enseignant-chercheur en informatique au Laboratoire d'Informatique de Grenoble, et à Grenoble INP - Université Grenoble Alpes. Ses travaux portent historiquement sur le choix social computationnel. Il s'intéresse depuis quelques années, dans le cadre de ses recherches et de son enseignement, aux enjeux environnementaux et sociétaux de l'informatique.

→ ***De l'eau dans les nuages***

Pascal BROSSET est *Managing Director* en charge du *Digital Manufacturing* chez Accenture, *leader* mondial du conseil et de l'intégration de systèmes. Il conseille des clients industriels sur l'adoption des technologies du *cloud computing* et de l'intelligence artificielle pour la conception, la validation et l'opération de systèmes de production. Diplômé en robotique industrielle de l'École Polytechnique Fédérale de Lausanne, il a supervisé la conception, la construction et l'opération d'usines intégrées puis occupé les

fonctions de responsable de la stratégie pour l'éditeur de logiciels de gestion SAP, puis de directeur de la technologie chez Schneider Electric.

→ *Systèmes cyber-physiques et jumeaux numériques, déclencheurs et conditions d'une adoption accélérée*

Aurélié BUGEAU est enseignante-chercheuse en informatique à l'Université de Bordeaux. Elle effectue ses recherches au LaBRI. Ses recherches portaient initialement sur le traitement et l'analyse d'images, notamment pour des applications de restauration d'images et de vidéos. Depuis 2020, elle s'intéresse au numérique responsable et notamment à la compréhension, la modélisation et l'enseignement des impacts environnementaux du numérique.

→ *De l'eau dans les nuages*

Laurent CLAVIER a obtenu son doctorat en traitement du signal à Telecom Bretagne (aujourd'hui IMT Atlantique) à Brest et le diplôme HDR de l'Université de Lille en 2009. Il est, depuis octobre 2011, Professeur à l'Institut Mines Télécoms (IMT Nord Europe) et responsable du groupe ARTS (Autonomous Resilient Systems). Il travaille aussi au sein de l'EMN (UMR CNRS 8520 - Institut d'Électronique de Microélectronique et de Nanotechnologie) et de l'IRCICA (USR CNRS 3380 - Institut de recherche sur les composants logiciels et matériels pour l'information et la communication avancée).

Ses activités de recherche concernent les communications numériques et la couche physique des réseaux sans fil pour l'Internet des Objets, plus particulièrement les réseaux de capteurs autonomes en énergie. Il s'intéresse particulièrement aux modèles d'interférence et à l'impact dans les réseaux sans fil ultra-denses. Il est président de l'action COST CA20120 INTERACT "Intelligence-Enabling Radio Communications for Seamless Inclusive Interactions".

→ *Du « sens » dans les réseaux*

Hervé DEBAR est actuellement Professeur de l'Institut Mines Télécom à Télécom SudParis, et directeur de la recherche de l'école. Il a été précédemment directeur du département réseaux de Télécom SudParis. Avant de rejoindre Télécom SudParis, il a été expert émérite en cybersécurité chez Orange, chercheur à IBM Zurich Research Laboratory, ingénieur-chercheur chez Dassault et CS. Il a un diplôme d'ingénieur de l'Institut National des Télécommunications (1990), un doctorat de l'Université Paris 6 (1993), et une Habilitation à Diriger des Recherches de l'Université de Caen (2006).

Hervé Debar travaille dans le domaine de la cybersécurité depuis plus de 30 ans, principalement sur la recherche et le développement d'outils de détection et de remédiation face aux attaques. Il est l'un des inventeurs du concept de "Security Information and Event Management", qui a pour but de donner aux opérateurs de cybersécurité l'intelligence situationnelle nécessaire pour évaluer en temps réel le risque lié aux activités des attaquants, et déployer les contre-mesures nécessaires. Dans ce domaine, il a été en particulier l'un des éditeurs du standard IDMEF (*Intrusion Detection Message Exchange Format*) et a déposé plusieurs brevets mondiaux dans le domaine. Il a publié plus d'une centaine d'articles scientifiques sur des sujets liés à la cybersécurité.

→ *La cybersécurité*

Emmanuel DOTARO est vice-président chez Thales et reconnu Thales Fellow expertise. Il est principalement en charge des sujets liés à la 6G au sein de l'organisation des directions techniques du groupe Thales.

Il obtient un Master en Informatique de l'Université de Versailles en 1996. Il effectue ensuite un doctorat au sein du laboratoire d'évaluation des performances de l'Institut National des Télécommunications pendant 3 ans, tout en occupant un poste d'enseignant à l'Université de Versailles.

Il rejoint en 1999 le laboratoire de Recherche et Innovation d'Alcatel à Marcoussis. Il devient ensuite directeur de recherche sur les réseaux chez Bell Labs, où il dirige successivement les services Packet Transport Infrastructure et Semantic and Autonomic Technologies.

Il intègre Thales en 2009 en tant que directeur de l'innovation pour les systèmes C4I et dirige jusqu'en 2022 les laboratoires ICT et Cybersécurité de Thales Secure Communication and Information Systems.

Il est auteur ou co-auteur de plus de 30 articles scientifiques et détient plus de 40 brevets dans le domaine des TIC. Il participe ou dirige divers projets de R&D collaboratifs européens ou nationaux et est membre de multiples comités techniques de conférences et revues internationales.

Parallèlement à ses activités actuelles de recherche et de gestion de l'innovation, il est impliqué dans diverses initiatives nationales et européennes structurantes dans les domaines des TIC et de la cybersécurité.

Ainsi, il est membre du Governing Board de l'association 6G-IA et impliqué dans des groupes de travail spécifiques (secteurs verticaux, politiques, sécurité...). Il est à l'origine de rapprochements et collaborations entre les communautés réseaux et sécurité. Il est également membre du comité de pilotage et du groupe d'experts de NetworldEurope. Il est éditeur du chapitre sur la sécurité de l'Agenda stratégique européen (SRIA) de recherche et d'innovation. Par ailleurs, il participe aux activités scientifiques, de certification et de recherche de l'association ECSO (communauté de la cybersécurité) et présidait le Cyber & Security Hub (plus de 260 membres) du *cluster* Systematic Paris-Région.

Il est enfin membre du conseil scientifique de l'Institut Mines Telecom et est président du comité des industriels du PEPR « Réseaux du futur » (France 2030) et est impliqué dans les Comités Stratégiques de Filières Industrie de sécurité et Infrastructure numériques.

→ ***Une perspective numérique européenne***

Le **Dr Denis DUTOIT** a rejoint le CEA en 2009 après avoir passé 17 ans chez STMicroelectronics et STEricsson en tant qu'architecte System-on-Chip. Il est actuellement impliqué dans de nombreux projets liés aux architectures multi-cœurs, hétérogènes et 3D pour le calcul haute performance (HPC) destiné aux supercalculateurs comme aux systèmes embarqués. Il a participé à plusieurs projets européens dans le domaine des télécommunications (Artist4G, BeFemto), des micro-serveurs (Euroserver) et du HPC (Mont-Blanc 1 et 2, Mont-Blanc 2020). Il a coordonné le projet H2020 ExaNoDe qui a abouti sur le développement d'un démonstrateur de nœud de calcul mixant une intégration 3D avec un System-in-Package.

Le Dr Denis Dutoit a également défini l'architecture de la « plateforme commune » de l'initiative de processeur européen (EPI), ossature de l'architecture des circuits issus du projet. Il est actuellement responsable du programme Calcul et Architecture Numérique du CEA-List. Il est l'auteur ou le coauteur de plus de 20 articles, dont certains articles invités ou tutoriels dans les conférences de premiers rangs : COFLER A., DRUILHE F., DUTOIT D. & HARRAND M. (2005), "A reprogrammable EDGE baseband and multimedia handset SoC with 6 Mb embedded DRAM", ISSCC. IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005, San Francisco, CA, USA ; DUTOIT D. *et al.* (2013), "A 0.9 pJ/bit, 12.8 GByte/s WideIO memory interface in a 3D-IC NoC-based MPSoC", Symposium on VLSI Technology, Kyoto, Japan ; DUTOIT D. *et al.* (2020), "How 3D integration technologies enable advanced compute node for Exascale-level High Performance Computing?", IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA.

Le Dr Denis Dutoit est titulaire d'un diplôme d'ingénieur de l'École Nationale Supérieure d'Électronique et de Radioélectricité de Grenoble, et d'un doctorat en traitement du signal de l'École Nationale Supérieure des Télécommunications de Paris.

→ ***Les composants pour les infrastructures numériques***

David GESBERT (Fellow, IEEE) est professeur et directeur d'Eurecom, Sophia Antipolis, France (www.eurecom.fr). Il a obtenu le doctorat de Télécom Paris en 1997. De 1997 à 1999, il a travaillé au Laboratoire des systèmes d'information de l'Université de Stanford. Il a été l'un des ingénieurs fondateurs d'Iospan Wireless Inc., une *spin-off* de Stanford, pionnière des technologies MIMO et OFDM pour les systèmes sans fil, qui fut vendue à Intel.

Avant de rejoindre Eurecom en 2004, il travaillait au département d'Informatique de l'Université d'Oslo. Il a publié environ 360 articles et 25 brevets, dont 7 ont remporté des prix de Best Papers par IEEE. Il a été co-président du programme technique d'ICC2017 et a été sélectionné comme chercheur « hautement cité » par Thomson-Reuters en informatique. Il est membre du conseil d'administration de l'OpenAirInterface (OAI), alliance logicielle mondiale pour l'*open source* dans la 5G, fondée par Eurecom. Il a été lauréat d'une ERC Advanced Grant dans le domaine des futurs réseaux. En 2020, il a également été sélectionné par l'Institut interdisciplinaire d'intelligence artificielle (3IA) pour une chaire dans le domaine de l'IA pour le futur IoT. En 2021, il a reçu le Grand prix de la Recherche conjointement de l'IMT et de l'Académie des Sciences pour ses travaux dans le domaine du MIMO.

→ ***La 6G : promesses et défis à l'horizon 2030***

Pierre GRONLIER, diplômé de l'EPITA en systèmes temps réel embarqués et diplômé de l'ENS Paris-Saclay en mathématiques appliquées, a démarré sa carrière avec une expertise en traitement du signal et en *machine learning*, qui l'a conduit à débiter en tant qu'ingénieur de recherche dans le domaine de la compression vidéo.

Puis, il a travaillé comme développeur chez Microsoft sur la couche multimédia temps réel de Skype/Teams (Lync). Plus tard, il s'est concentré sur la construction et l'expansion des offres de services *cloud*, chez OVHcloud en France en tant qu'architecte solution.

En 2021, il a rejoint Gaia-X en tant que Chief Technology Officer où il a mis en place et développé toutes les premières versions des services fédérés. Il a récemment élargi ses fonctions en tant que Chief Innovation Officer de Gaia-X.

En 2024, il a cofondé CloudDataEngine, entreprise logicielle spécialisée dans la numérisation, la traçabilité et la composition des certificats de conformité, tel que CSRD.

→ ***Les apports de Gaia-X***

Fabrice GUILLEMIN est Ingénieur général des Mines. Il est diplômé de l'École Polytechnique (X84) en 1987 et de Telecom Paris en 1989. Il a défendu une thèse à l'Université de Rennes 1 en 1992 et une thèse d'habilitation à diriger des recherches à l'Université Pierre et Marie Curie en 1999. Il a rejoint les laboratoires de recherche d'Orange à Lannion en 1989 (à l'époque le Centre National d'Étude des Télécommunications de France Télécom). Il a occupé plusieurs fonctions au sein de la Recherche à Orange. Il est actuellement responsable d'un programme de recherche sur l'architecture des réseaux virtualisés. Il est également Senior Expert de la Communauté "Network of Future" (NoF) à Orange.

→ ***Les réseaux virtualisés : promesses et enjeux***

Francis JUTAND a travaillé dans le domaine de la formation et de la recherche en numérique à Télécom Paris, puis a dirigé Télécom Bretagne. Il a été directeur scientifique du Centre de R&D d'Orange, du département numérique du CNRS, de l'Institut Télécom puis de l'IMT dont il a été ensuite directeur général adjoint en charge de la stratégie et du développement. Il a présidé le Conseil scientifique numérique de l'ANR. Il a créé le pôle de compétitivité Cap Digital et a été à la création de l'Institut Carnot Télécom et Société Numérique. Il a été membre du Conseil national du numérique.

Prospectiviste du numérique, il s'intéresse aux impacts scientifiques, techniques, économiques, d'usage et éthique. Il a été éditeur du livre « Métamorphose numérique vers

une société de la coopération et de la connaissance » et auteurs dans plusieurs ouvrages de prospectives.

Impliqué dans le numérique, les infrastructures, les données, l'intelligence artificielle, les usages, il est un des fondateurs de la chaire sur « les données personnelles et les identités numériques » et de la chaire sur « l'Économie des Communs de Données » de l'IMT. Il est membre du conseil scientifique de OnePoint, administrateur de la Société Française de Prospective, associé au *think tank* Digital New Deal, et a été nommé membre associé du Conseil général de l'Économie.

→ **Introduction**

→ **Des services réseaux aux plateformes du numérique et aux services multisectoriels**

→ **Le rôle de l'IA dans l'évolution des infrastructures numériques du futur**

Daniel KOFMAN, Professeur à Telecom-Paris, est co-fondateur et directeur exécutif du LINCS, un laboratoire académique-industriel de recherche et d'innovation sur les réseaux de communication, les systèmes et les services du futur, financé par l'Institut Mines Telecom, Inria, Sorbonne Université, Nokia et SystemX.

Il est également co-fondateur et directeur exécutif de ICT4V, un centre technologique créé pour exploiter les opportunités des technologies de l'information en tant que moteur majeur de l'innovation dans tous les secteurs verticaux.

Daniel Kofman est co-directeur du Programme national français « PEPR Future Networks », bras de la recherche de la stratégie nationale française dans le domaine.

Il est impliqué dans l'écosystème des *start-ups* depuis plus de 20 ans, notamment étant lui-même co-fondateur de 4 *start-ups*.

Il intervient en tant qu'expert et consultant auprès de diverses entreprises (opérateurs télécoms, constructeurs, *venture capitals*), ainsi qu'auprès d'institutions nationales et internationales.

Il a été CTO puis Fellow-Strategic Advisor de RAD Data Communications, ainsi que membre de son comité de stratégie (CEO, 5 VPs et CTO) et membre du Comité scientifique du Parlement français (OPECST).

Ses principaux sujets de Recherche et d'Innovation sont les réseaux de communication du futur, la convergence réseau-*cloud*, la transition numérique, les *smart-grids* et la transition énergétique.

Il est l'auteur de plusieurs documents de vision et de positionnement pour l'industrie et les organismes publics, ainsi que de plus de 100 articles scientifiques.

→ **Introduction**

→ **Des services réseaux aux plateformes du numérique et aux services multisectoriels**

→ **Le rôle de l'IA dans l'évolution des infrastructures numériques du futur**

Marios KOUNTOURIS a obtenu son diplôme en génie électrique et informatique de l'École Polytechnique d'Athènes (NTUA), Grèce, en 2002 et son Master et doctorat de Télécom ParisTech, France, en 2004 et 2008, respectivement. Il est actuellement Professeur au département des systèmes de communication, EURECOM, France.

Avant cela, il a occupé des postes à CentraleSupélec, au Huawei Paris Research Center, en France, à l'Université du Texas à Austin, aux États-Unis, et à l'Université de Yonsei, en Corée du Sud. Il a été rédacteur en chef d'IEEE Transactions on Wireless Communications, IEEE Transactions on Signal Processing et IEEE Wireless Communication Letters. Il est récipiendaire d'une Consolidator Grant du conseil européen de la recherche (ERC) en 2020 sur la communication sémantique orientée vers un objectif. Il a reçu plusieurs prix et distinctions, dont la médaille Blondel 2022 et le prix IEEE ComSoc Outstanding Young Researcher Award 2013 pour la région EMEA. Il est fellow de l'IEEE, de l'AAIA et de l'AIHA.

→ **Du « sens » dans les réseaux**

Dimitri KTÉNAS a obtenu le diplôme d'ingénieur en électronique et traitement du signal de l'École Nationale Supérieure d'Électronique et de Radioélectricité de Grenoble (ENSERG) en 2001. Depuis, il travaille au CEA-Leti à Grenoble (France). Ses principaux domaines de recherche actuels sont l'optimisation PHY, MAC et inter-couches pour la 6G. Il a participé à plusieurs projets européens et a été le coordinateur du projet français OPUS qui portait sur l'optimisation du LTE. De 2010 à 2015, il a dirigé le laboratoire Études des Systèmes de Communication sans Fil au sein du CEA-Leti, dont l'objectif était l'étude de traitements en bande de base et au niveau de la couche MAC pour les systèmes sans fil. En 2016, il a pris la direction du laboratoire Systèmes sans Fil Haut Débit au sein du CEA-Leti, en charge des études algorithmiques et de l'implémentation HW/SW du traitement du signal numérique et des protocoles pour les systèmes 5G et LiFi.

En mars 2018, il a été nommé chef du service des technologies sans fil, se concentrant sur le B5G/6G, l'IoT industriel, les communications optiques sans fil et les capteurs RF, de la bande de base aux couches réseau, en incluant la modélisation du canal de propagation, la conception d'antennes et la conception de circuits intégrés RF. Depuis mai 2023, il est également co-directeur du PEPR Réseaux du futur, qui est un programme français visant à concevoir à l'horizon 2030 les futurs réseaux et systèmes en réseau. Il a publié plus de 70 articles scientifiques dans des revues internationales et des actes de conférence, ainsi que 5 chapitres de livres, et est l'inventeur ou le co-inventeur principal de 13 brevets.

→ ***Les composants pour les infrastructures numériques***

Laura LÉTOURNEAU est ingénieure du Corps des Mines. Elle a d'abord travaillé dans le photovoltaïque chez SunPower avant de rejoindre Solvay comme directrice de projet Biotech au Brésil. Elle a ensuite co-écrit l'ouvrage « Ubérisons l'État ! Avant que d'autres ne s'en chargent » (éditions Armand Colin, préface de Xavier Niel) qui prône l'amélioration des services publics en s'appuyant sur le modèle de plateforme publique. Elle a mis en œuvre cette vision à l'Arcep (Autorité de régulation des communications électroniques et des postes) en tant que cheffe de l'unité « Internet ouvert », puis comme déléguée ministérielle au numérique en santé auprès des ministres chargés de la Santé, notamment pendant la crise du Covid-19.

En parallèle, elle a co-fondé le réseau informel « À Cœur Publics », constitué d'agents qui se serrent les coudes pour tenter d'améliorer le fonctionnement de l'administration. Elle a ensuite été missionnée par la Première ministre pour répliquer la méthode de plateforme publique à la planification écologique, ce qui a donné lieu à la feuille de route « Numérique et Données » de France Nation Verte, ainsi que pour tirer des enseignements de la transformation numérique publique.

Laura Létourneau figure dans le Palmarès 2024 Cercle de Giverny x Le Point des 50 acteurs de la transformation écologique et sociale de moins de 40 ans. Elle fait partie du réseau des « Femmes de Tech » de l'Académie des Technologies et des French-American Young Leaders. Elle s'est vue remettre la médaille des Chevaliers de l'Ordre du Mérite.

→ ***Plaidoyer pour les grandes oubliées : les infrastructures publiques de partage de données***

Éric MONCEYRON est Ingénieur des travaux publics de l'État (1986), titulaire d'un DEA informatique et automatique appliquées de l'Université Claude Bernard Lyon 1 associée à l'École des Mines de Saint-Étienne (1986), et d'un doctorat en contrôle des systèmes (intelligence artificielle) de l'Université de Technologie de Compiègne (1991). Auditeur de l'Institut des Hautes Études d'Aménagement des Territoires (cycle territoires et mobilités - 2020).

Il est ingénieur en chef de la fonction publique territoriale et, au cours de sa carrière, a dirigé des projets multi-partenariaux dans l'innovation numérique, le transport et la mobilité, l'aménagement, les infrastructures et la transition énergétique, et des politiques publiques d'aménagement du territoire et d'urbanisme.

Il a exercé les responsabilités de directeur général adjoint au sein de deux communautés d'agglomération jusqu'en 2011, après avoir piloté des équipes pluridisciplinaires et des grands travaux au sein de services du ministère de l'Équipement.

Il a coordonné pour Bordeaux Métropole la mission pour l'organisation du congrès mondial des systèmes de transports intelligents en 2015, en lien étroit avec l'écosystème.

→ ***Les systèmes de transports intelligents coopératifs au service des politiques de mobilité urbaine***

Guillaume MOREAU a obtenu son doctorat à l'Université de Rennes I en 1998, puis son habilitation à diriger des recherches de l'université de Nantes en 2009. Il débute sa carrière en 1999 comme ingénieur de recherche à l'École des Mines de Paris, exerce ensuite comme maître de conférences puis Professeur des universités en informatique à l'École Centrale de Nantes. En 2020, il rejoint IMT Atlantique comme directeur délégué à la recherche et à l'innovation.

Ses recherches concernent la réalité virtuelle et la réalité augmentée et plus spécifiquement en ce moment la perception dans les environnements virtuels et augmentée. Il collabore avec les équipes d'Inria Rennes, de l'Université de Tokyo et de l'Université d'Australie Méridionale. Il a participé à de nombreux projets d'introduction de ces technologies dans l'industrie manufacturière. Il est également membre du comité de pilotage de la conférence de référence dans le domaine, IEEE International Symposium on Mixed and Augmented Reality. Il est membre de l'Advisory Council du Reality Labs de Meta. Il a été également membre ou président de comités d'évaluation de l'Agence Nationale de la Recherche.

→ ***Métavers : au-delà des « casques de réalité virtuelle »***

Anne-Cécile ORGERIE est chercheuse au CNRS à l'IRISA à Rennes. Elle fait partie de l'équipe Magellan qui travaille sur les systèmes distribués à large échelle. Ses travaux de recherche portent sur la mesure, la modélisation, la simulation et l'amélioration de l'efficacité énergétique des systèmes distribués. Elle dirige actuellement le groupement de recherche et de service EcoInfo qui étudie les impacts environnementaux et sociétaux du numérique.

→ ***De l'eau dans les nuages***

Sébastien PICARDAT est, depuis 2020, le directeur général d'Agdatahub, l'infrastructure française de partage de données agricoles. Avec plus de 20 ans d'expérience dans le secteur agroalimentaire, il est passionné par l'exploitation des données et de la technologie en tant que sources de valeur et d'impact pour l'industrie agricole et alimentaire en France et en Europe.

Expert en gestion de projet, en conseil stratégique, en gestion du changement, en stratégie numérique, en développement de réseaux et de partenariats, et en représentation institutionnelle, il a développé et mis en œuvre avec succès des projets complexes et transversaux, tels que AgriDataSpace, le futur espace européen de données agricoles. Il s'engage également à favoriser la collaboration et le dialogue entre les différentes parties prenantes et les acteurs de l'écosystème agroalimentaire, depuis les agriculteurs et les coopératives jusqu'aux autorités publiques et aux organisations interprofessionnelles.

→ ***Apport de l'infrastructure de partage de données Agdatahub dans le secteur agri-agro***

Emmanuel PUIG, dirige actuellement le département R&D d'Amarisoft, éditeur logiciel spécialiste des technologies mobiles 4G et 5G.

Diplômé de l'École Centrale Paris, après avoir participé au logiciel VLC, il a fait ses classes au sein de l'industrie des télécoms dans la télévision numérique sur IP puis a rejoint Amarisoft pour mettre en place l'équipe de développement et de recherche.

→ ***La softwarization des réseaux et son impact technico-économique***

Sophie QUINTON est chercheuse à l'Inria Grenoble. Ses travaux de recherche portaient jusqu'à 2018 sur la vérification formelle de systèmes embarqués. Elle étudie désormais les enjeux de durabilité liés au numérique, et en particulier la matérialité et les valeurs sous-jacentes aux infrastructures et aux systèmes informatiques ainsi qu'à la place du numérique dans des stratégies de décroissance.

→ *De l'eau dans les nuages*

Claudine RABILLARD est directrice déléguée Territoires d'Enedis, en charge du développement des Solutions pour la Transition énergétique pour les collectivités locales, des partenariats au service de la transition énergétique des territoires et de l'animation de la filière territoriale d'Enedis. Elle était auparavant en charge du marketing et du développement des offres de services énergétiques du groupe EDF à l'international. Elle a également 10 ans d'expérience dans la communication notamment au service des projets de transformation du groupe et a exercé diverses missions dans les fonctions RH *corporate* d'EDF.

Claudine Rabillard a commencé sa carrière à la direction R&D d'EDF en tant qu'ingénieur-chercheur. Elle est titulaire d'un doctorat en analyse numérique appliquée à la mécanique des fluides de l'Université Lyon I-École Centrale de Lyon.

→ *Les enjeux du numérique dans la distribution d'électricité*

Anne-Sophie TAILLANDIER, diplômée de CentraleSupélec et titulaire d'une thèse en *Machine Learning* de l'ENS Paris Saclay, a une large expérience en édition logicielle et dans l'industrie. Elle a occupé différentes fonctions de direction (développement produits et programmes, R&D) pendant 10 ans chez Dassault Systèmes, avant de prendre la direction des développements en authentification dédiée aux applications bancaires.

Après quelques années chez LTU Technologies comme directrice de la technologie, responsable de l'activité technique et des développements (logiciel de reconnaissance d'images), elle est depuis 2015 directrice générale de TeraLab, filiale de l'IMT, acteur majeur de l'accélération et de l'adoption des technologies *big data* et IA dans les différents secteurs de l'économie. Elle a participé à la création de Gaia X AISBL dont l'IMT est devenu membre fondateur. Elle est au conseil d'administration de Cap Digital et de CentraleSupélec. Elle a été élue membre de l'Académie des Technologies en 2022.

→ *Les apports de Gaia-X*

Claire WAAST-RICHARD est directrice Data et Numérique d'Enedis. Elle a débuté sa carrière à IBM en 1990, en R&D dans le domaine de l'intelligence artificielle. Elle a dirigé plusieurs équipes en Europe, aux États-Unis et a travaillé en étroite collaboration avec la Chine. En 2005, elle rejoint la direction R&D d'EDF pour développer des solutions informatiques et mathématiques appliquées à l'analyse du comportement des clients.

En 2009, elle devient directrice de Cabinet R&D et, en 2011, directrice des systèmes d'information en charge également du programme R&D des technologies de l'information. En 2016, elle rejoint Enedis en tant que directrice adjointe des systèmes d'information. Fin 2021, elle prend le pilotage de la transformation *data-driven* de l'entreprise. Elle est titulaire d'un doctorat en Intelligence artificielle et Traitement du Signal, ingénieure diplômée de Telecom Paris.

→ *Les enjeux du numérique dans la distribution d'électricité*