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 2022[1].

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.

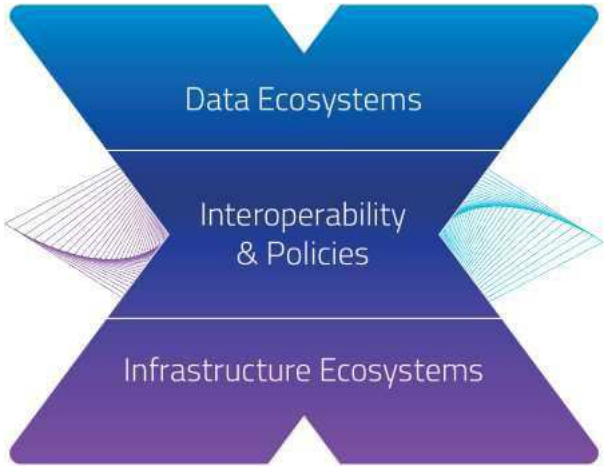

Figure 1 : Gaia-X concerne les écosystèmes d'infrastructure et de données (Source : Gaia-X — Hub France).

L'initiative doit permettre de porter les valeurs européennes de transparence, d'autodétermination, 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 numériques 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-X[2] 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 réplicabilité, 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 du 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.

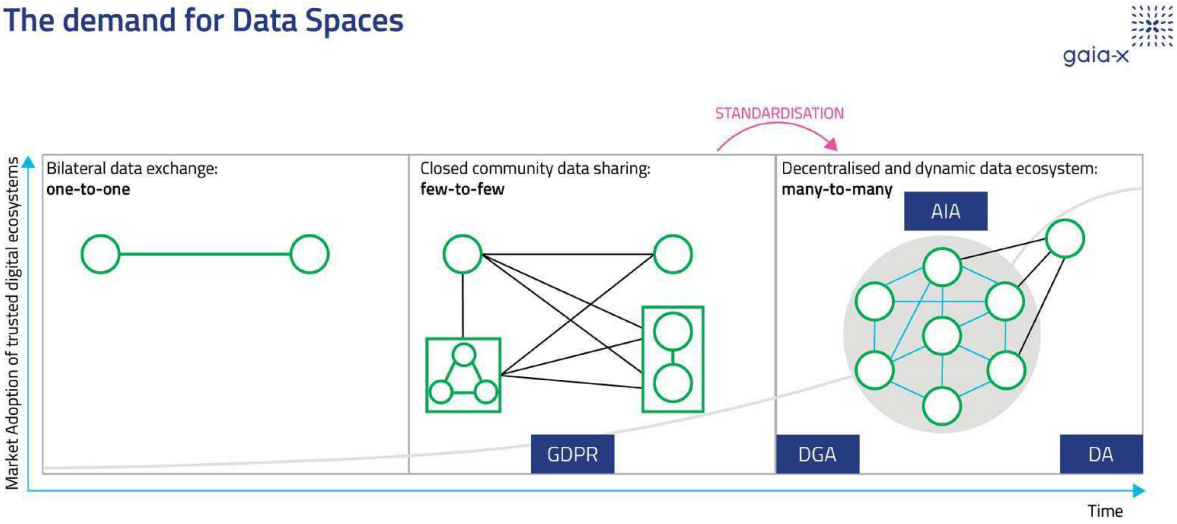

Figure 2 : La décentralisation est essentielle pour le passage à l'échelle (Source : Data Spaces Business Alliance).

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 déclaration 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 Framework[3].

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.

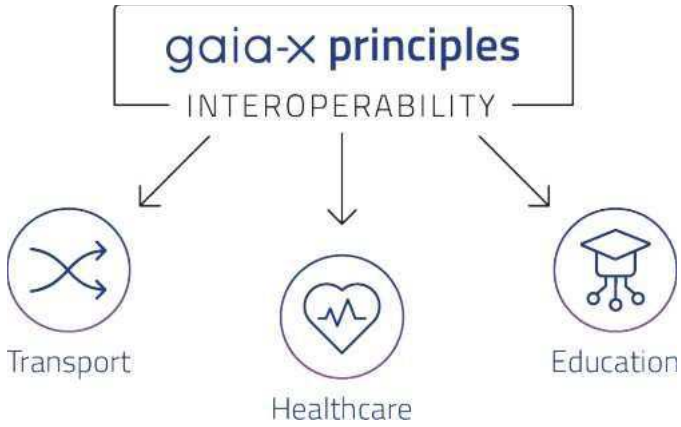

Figure 3 : Interopérabilité entre
les différents domaines (Source : Gaia-X).

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 compliance[4].

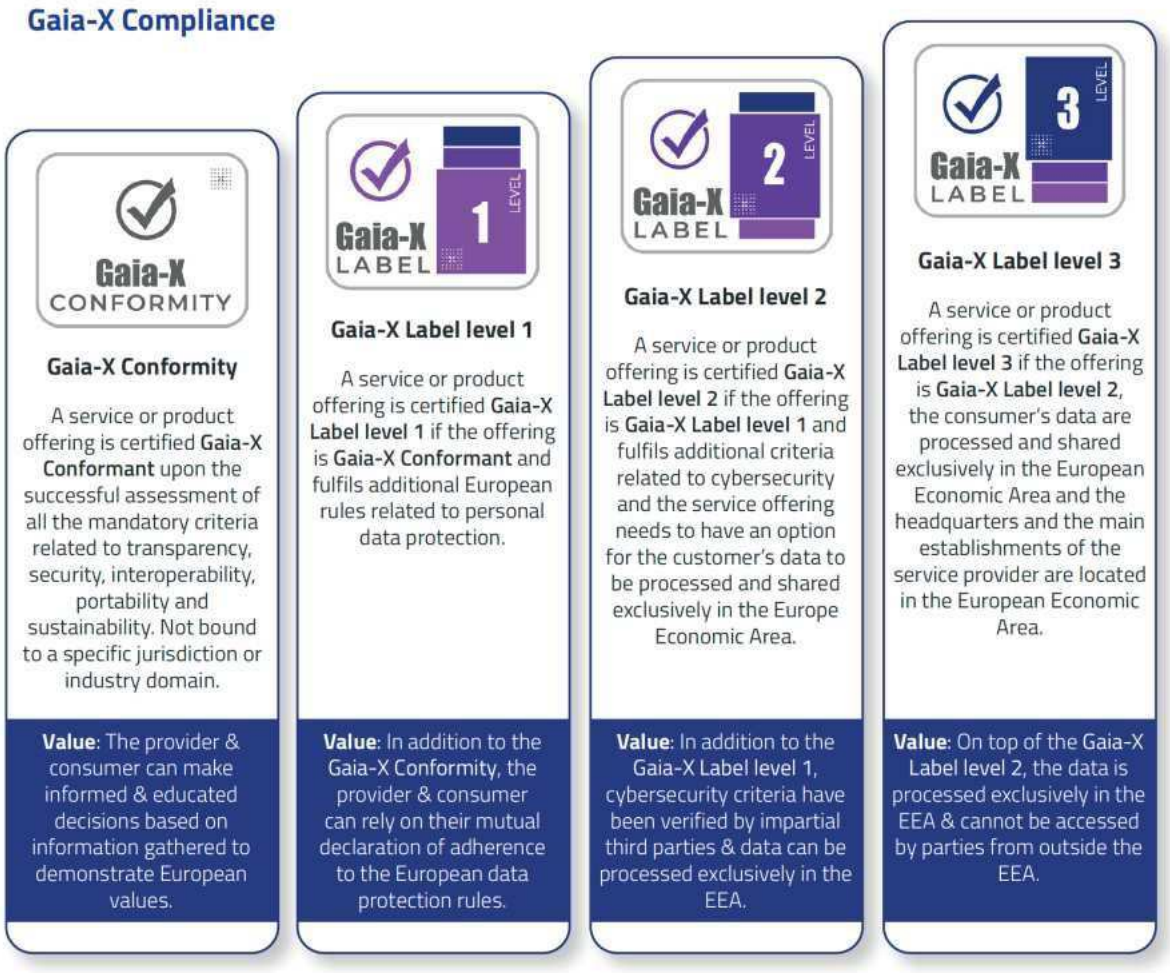

Figure 4 : Les niveaux de confiance de Gaia-X (Source : Gaia-X).

[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/