

# Le *Big data* et l'*Open data* au service des collectivités

Par Pascal SOKOLOFF

Fédération nationale des collectivités concédantes et régies (FNCCR)

La FNCCR (Fédération nationale des collectivités concédantes et régies), qui accompagne ses adhérents (collectivités locales et groupements de collectivités) dans la mise en œuvre de leurs compétences en matière de réseaux publics (énergie, eau et assainissement, numérique) considère que le phénomène de la transition numérique impacte considérablement les processus dans ce domaine.

C'est pourquoi elle a entrepris de rédiger un Livre blanc du numérique, pour faire partager son point de vue sur cette mutation irréversible. Cet article est directement issu de nos réflexions et d'analyses collectives produites dans un contexte participatif. Nous invitons le lecteur à consulter ledit Livre blanc (<http://www.fnccr.asso.fr/wp-content/uploads/2016/06/livre-blanc-numerique-FNCCR.pdf>), dans lequel sont développés de façon plus détaillée les enjeux et les perspectives de l'utilisation des données dans la gestion publique.

**D**ans une économie de plus en plus mondialisée et de plus en plus affectée par le numérique, la donnée devient un élément de valeur ajoutée, et donc de richesse, de plus en plus visible, et ce d'autant plus qu'elle circule sans contrainte et de manière quasi instantanée. La conséquence de cet état de fait se mesure en chiffres : en 2013, ce sont 100 milliards de messages qui étaient échangés chaque jour.

Mais ce chiffre en soi n'aurait pas de valeur si ces données n'étaient pas captées et exploitées par des entreprises capables d'en retirer une information plus riche et plus exhaustive sur les pratiques de milliards de consommateurs, afin d'en comprendre les comportements et d'ajuster leur stratégie *marketing* en conséquence. Des entreprises numériques comme Google, Amazon ou Uber sont capables d'organiser l'information à l'échelle mondiale à partir des flux générés par les utilisateurs de leurs services (moteurs de recherche et navigateur pour Google, historique des achats ou des consultations d'objets culturels pour Amazon...), et bien d'autres encore, sont désormais capables de capter une partie de la valeur de ces données et de générer, grâce à elles, de nouvelles richesses.

Mais à l'exception de rares réussites, notre pays peine à faire émerger des champions numériques. Les administrations et, de manière générale, les acteurs publics sont restés, en France, longtemps à l'écart de ces différents aspects de la transformation numérique. Aujourd'hui, ils commencent à l'accompagner – à défaut de la précéder – pour tenter de capter une partie de cette valeur. Nos administrations et nos entreprises doivent être davantage en

phase avec l'économie numérique : elles doivent numériser leur organisation, collecter plus de données et être davantage présentes en ligne afin de se rapprocher des citoyens ou de leurs clients.

En France, le très haut débit peut être un vecteur de croissance et d'accélération de cette transition numérique (on constate en effet que les territoires non connectés se marginalisent de plus en plus rapidement). Mais il n'est qu'un véhicule, qu'une autoroute, que les voyageurs pourront emprunter : il n'est pas à même de transformer, à lui seul, les façons de produire et de consommer. Or, dans le type d'échanges qu'il va permettre, l'enjeu pour les territoires est bien de pouvoir disposer d'un écosystème à la fois administratif et entrepreneurial qui soit susceptible de prendre appui sur la Toile pour pouvoir se démultiplier. Tout d'abord à l'échelle de la ville et des territoires, puis à celle du pays afin de lui permettre de retrouver une partie de sa souveraineté.

Dans le cadre de la discussion du projet de loi pour une République numérique, l'État réfléchit aux missions qui pourraient incomber à un futur service public de la donnée. Sans préjuger de son positionnement définitif, il faut reconnaître qu'il demeure imprécis. Le texte de loi ne mentionne pas explicitement la création d'un tel service public, se contentant d'indiquer que « la mise à disposition de données de référence en vue de faciliter leur réutilisation » constitue une mission de service public relevant de l'État. Toutes les administrations – l'État, les collectivités territoriales et les autres personnes morales relevant du droit public – devront concourir à cette mission.

Les données dites de référence sont des données publiques qui « constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes ». Ce sont des données qui sont « réutilisées fréquemment par des personnes publiques et privées autres que l'administration qui les détient » et qui, par conséquent, nécessitent d'être « mises à disposition avec un niveau élevé de qualité ».

C'est (en principe) un décret en Conseil d'État qui précisera les modalités de participation et de coordination des différentes administrations dans l'exercice de cette mission, ainsi que les critères de qualité que celles-ci devront respecter. Ce décret dressera également la liste des données de référence et désignera les administrations responsables de leur production et de leur mise à disposition. L'État aura donc la charge de définir et d'identifier les données de référence, d'harmoniser le cadre et les conditions de leur production et d'en fixer des normes de qualité et les modalités de diffusion.

Le Secrétariat général pour la modernisation de l'action publique (SGMAP) sera très vraisemblablement chargé d'assurer cette mission de service public, ce qui pose la question du rôle des collectivités territoriales qui, en raison de leur position et de leurs missions, figureront parmi les principaux fournisseurs de données de référence. Comment lesdites collectivités y seront-elles associées et comment la charge du financement sera-t-elle répartie ? Si quelques amendements déposés par des parlementaires ont bien proposé que les collectivités territoriales partagent avec l'État la gouvernance de cette mission, ils ont tous été rejetés, le gouvernement renvoyant le débat aux discussions et aux négociations préparatoires prévues dans le cadre de la rédaction du futur décret.

### La donnée, source d'innovation et de développement de nouveaux services

Si l'utilisation de données à des fins de traitement et d'analyse est inhérente à tout pilotage d'organisation, le changement d'échelle provoqué par Internet en matière de données a permis le développement de nouveaux services grâce à la collecte et à l'analyse en temps réel de données massives, et grâce à des croisements de données inédits (entre données *Open Data*/données non structurées/données de géolocalisation, etc.).

C'est ce que l'on appelle le *data mining* (littéralement, le « forage des données ») – ou l'exploration massive de données qui permet d'analyser les informations issues de bases de données et d'établir entre celles-ci d'éventuelles corrélations significatives et utilisables. Si ces fonctions sont aujourd'hui très largement utilisées dans le domaine de la « connaissance client » afin d'apporter une aide à la décision performante dans les stratégies *marketing*, elles pourraient l'être davantage dans la connaissance des besoins des citoyens (en préservant leur anonymat). C'est ce que nous appelons à la FNCCR le *Big data territorial*.

Le *data mining* (qui est pratiqué depuis les années 1980) s'appuie désormais sur l'ensemble des outils et des méthodes d'extraction des connaissances actuelles comme

le *crawling* ou le *scraping* de données <sup>(1)</sup>, ou encore le *machine learning* <sup>(2)</sup>. Depuis l'invention du papier et des registres (et même bien avant), les collectivités ont toujours cherché à classer et à conserver les données concernant leurs habitants ou la distribution des parcelles de leur territoire. Aujourd'hui, l'augmentation du nombre de données et, en parallèle, la montée en puissance des ordinateurs ont installé la donnée au cœur des processus de gestion. Aussi le secteur public peut-il avoir recours au numérique pour atteindre trois types de finalité :

- la recherche d'économies, grâce à la simplification des procédures et à une mutualisation des ressources et des compétences,
- la recherche de la performance, pour fournir un service public plus efficace et plus personnalisé (grâce à la collecte et au traitement de nombreuses données personnelles),
- l'activation de la citoyenneté, en créant (par exemple) des outils d'entraide de proximité, de *crowdsourcing* (production participative), de consommation collaborative et de participation citoyenne.

À travers ces trois finalités, on note la place importante prise par la gestion des données dans l'administration publique et, surtout, dans sa propre révolution. En effet, on assiste à une diversification sans précédent des sources et des modes de production des services publics en lien avec l'informatisation des systèmes d'information (SI) – notamment avec l'introduction de la messagerie électronique –, la numérisation des services publics (avec l'adoption par de nombreuses villes de la carte de vie quotidienne – CVQ), l'Internet des objets et le *monitoring* urbain (capteurs, vêtements connectés, coproduction avec les usagers, analyse des comportements de mobilité, prédiction des probabilités d'embauche par secteur et services personnalisés).

Cela s'accompagne de nouvelles logiques pour l'action publique qui s'appuient désormais sur des politiques « consommatrices » de données et sur la coopération et l'échange de données entre administrations (notamment dans le cadre de programmes, tels que *Dites-le-nous une fois* (le programme qui vise à simplifier la vie administrative des entreprises en réduisant le nombre des informations redondantes qui leur sont demandées) ou *France-Connect* (qui vise à simplifier la relation des usagers avec l'ensemble des administrations). Ces nouveaux usages concernent les problématiques de gestion à la fois d'un territoire et de la relation citoyenne. Mais la valeur ajoutée de ces nouveaux usages et services est intrinsèquement liée au partage et à la réutilisation de données. En effet, c'est aussi la circulation des données qui permet de créer de la valeur.

Nous pensons donc que les collectivités locales doivent s'approprier les outils adaptés à cette révolution.

(1) Le *crawling* et le *scraping* sont deux techniques différentes d'acquisition automatique des données sur le Web, qui permettent de convertir celles-ci directement en données structurées.

(2) Techniques permettant d'obtenir des analyses prédictives à partir de formes d'apprentissage de la machine via les données.

## Des missions locales de service public pour gérer des données à la maille régionale

Si la question des données de référence est cruciale, en termes de valeur ajoutée et de service, elle ne constitue pour les collectivités territoriales que l'une des facettes de la gouvernance de la donnée. Au-delà se posent des questions d'organisation des entités publiques locales impliquées dans cette gouvernance, d'élargissement de l'ouverture des données publiques et des données des entreprises (lorsqu'elles sont d'intérêt général), de prise en compte des inégalités territoriales (notamment par rapport à la capacité des collectivités à financer et à mettre en œuvre les moyens de production et de diffusion nécessaires à cette ouverture) et d'appropriation des *data sciences* en matière de développement des territoires.

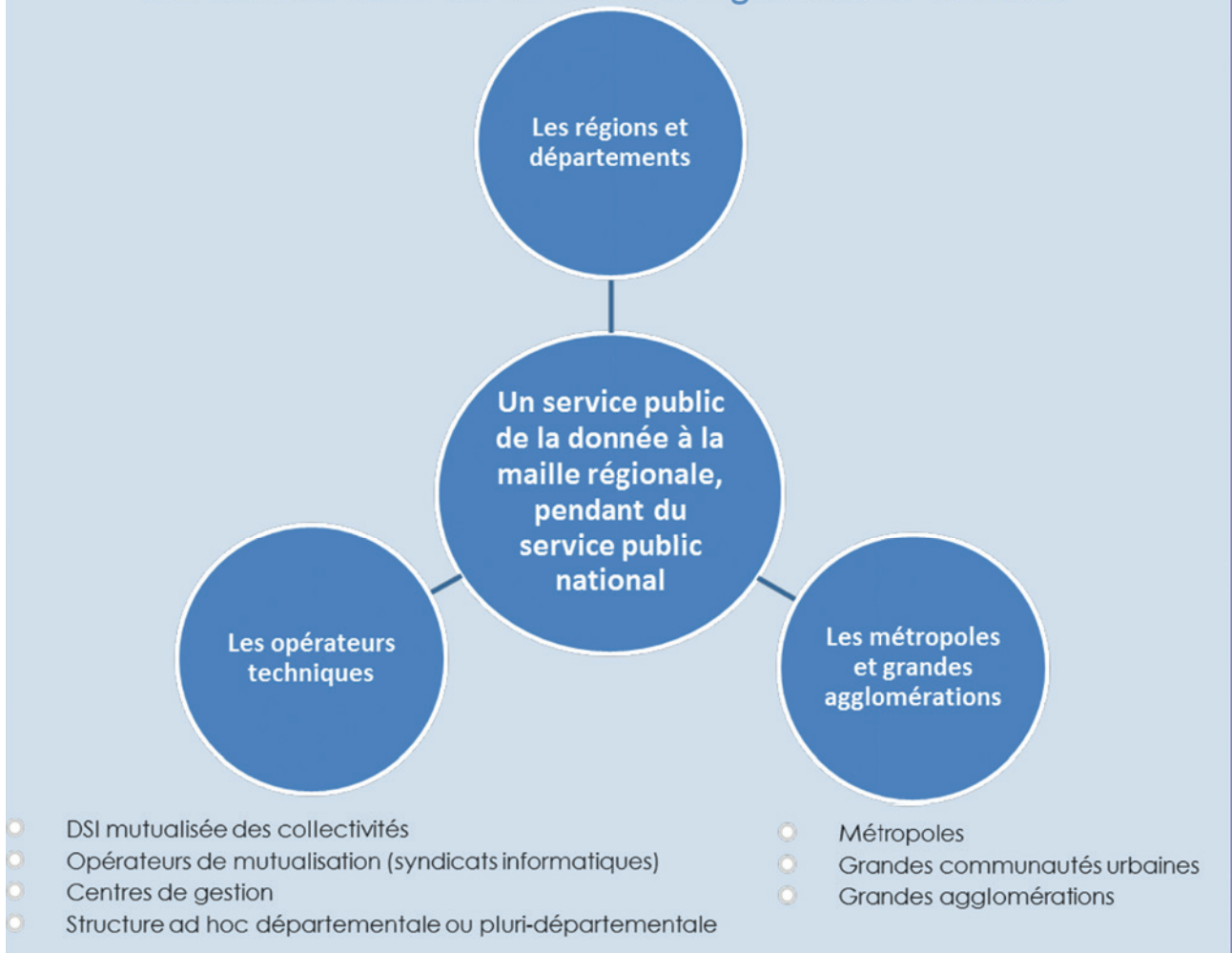
Tous ces sujets induisent de nouvelles formes de collaboration, d'échange, de mutualisation de projets, et donc de coordination dans la mise en œuvre d'infrastructures de données publiques locales. Une coordination et des moyens communs qui pourraient être pilotés et gérés par un service local de la donnée. Cette entité devrait être légère, flexible, à géométrie variable et être au service du

secteur public et des acteurs économiques locaux. Complémentaire à ce que réalisent déjà certaines collectivités (les métropoles) par elles-mêmes, ce service local de la donnée pourrait intervenir plutôt à l'échelle régionale pour garantir la cohérence des actions et disposer de leviers d'action suffisants. Mais il serait tenu à une obligation de neutralité et au respect de règles de confidentialité équivalentes à celles de tiers de confiance afin d'accompagner si nécessaire une filière économique dans sa stratégie de redéploiement.

L'hétérogénéité des approches et des contextes territoriaux induit une grande flexibilité dans la structuration et l'organisation de cette mission. Néanmoins, cette dernière devrait rassembler, si possible, quatre familles d'acteurs, dans une configuration qui reste à déterminer :

- les régions et les départements,
- les métropoles et les grandes agglomérations,
- les opérateurs publics techniques qui, présents sur le territoire, accompagnent notamment les plus petites collectivités (syndicats intercommunaux ou mixtes, structures de mutualisation dans le domaine du numérique),
- les représentants des acteurs économiques locaux.

### Un service local de la donnée à géométrie variable



Le service en lui-même se situerait plutôt à la maille régionale, mais la région (tout comme les départements), bien que largement impliquée, n'en serait pas le pilote : une structure *ad hoc* créée spécialement pourrait endosser ce rôle.

La maille régionale semble être aujourd'hui l'échelle pertinente pour installer cette mission locale de service public de la donnée. Elle offre en effet une visibilité étendue surtout après le redécoupage récent du territoire national en 13 unités régionales. À ce niveau s'exerce déjà une coordination des données géographiques au travers des Centres régionaux de l'information géographique (CRIGE) et de certaines structures de mutualisation publiques (comme les groupements d'intérêt public Aménagement du territoire et gestion des risques – GIP ATGERI). La loi NOTRe (sur la nouvelle organisation territoriale de la République) a d'ailleurs officialisé cette coordination en la confiant aux régions, à charge pour elles de constituer une plateforme de services numériques et de l'animer pour en favoriser l'accès et la réutilisation des données. La proximité et l'implication des instances régionales en tant que chefs de file de la stratégie d'aménagement du territoire et du développement économique seraient également un atout, sachant qu'il s'y développera progressivement une culture de la collaboration et de la concertation avec l'ensemble des autorités publiques locales <sup>(3)</sup>. Outre la gouvernance de la donnée, la structure considérée devra conduire à l'échelle régionale trois autres missions principales : une mission de sensibilisation et d'accompagnement, la dynamisation du développement du territoire et, enfin, le développement des coopérations avec les autres collectivités et avec le privé.

### Première mission : la gouvernance de la donnée

Elle s'articule autour de quatre axes :

- création d'un schéma des données d'intérêt général sur le territoire, d'un magasin d'API (interfaces de programmation applicative) et d'un annuaire de métadonnées et déclinaison au niveau régional des outils susceptibles de valoriser le territoire régional,
- mise en place d'une infrastructure de données publiques incluant la création d'une plateforme et d'un volet *Open Data*,
- constitution de modèles de données métiers,
- veille et partage de bonnes pratiques sur l'utilisation et la réutilisation des données de référence et sur les API diffusées par l'État et les autres niveaux territoriaux.

### Deuxième mission : sensibilisation et accompagnement

Une mission qui s'adresse aussi bien aux acteurs et aux territoires qu'aux entreprises.

#### Mobilisation des acteurs et des territoires

- production et diffusion de bonnes pratiques et de retours d'expériences dans les organisations privées et publiques,
- développement de nouvelles filières de formation,
- promotion de l'innovation.

### Accompagnement du monde économique à travers trois axes

- la veille économique et les entrepôts de données partagées,
- l'organisation des filières économiques qui en expriment le besoin,
- les appels à projets sur des actions mobilisant les *data sciences*.

### Troisième mission : la dynamisation du développement du territoire

Une mission qui se concrétise au travers :

- de la création de tableaux de bord économiques régionaux dynamiques (en partenariat avec l'INSEE),
- d'applications mobiles de données dynamiques permettant à leurs utilisateurs de créer leur propre tableau de bord de vie quotidienne (transports, événementiels...),
- de tableaux de bord e-gov de l'usager, du touriste ou du voyageur en déplacement,
- des *data sciences*, en s'associant avec les *French Tech* locales, les laboratoires de recherche, les universités et les écoles d'ingénieurs, et en développant des services d'intelligence économique.

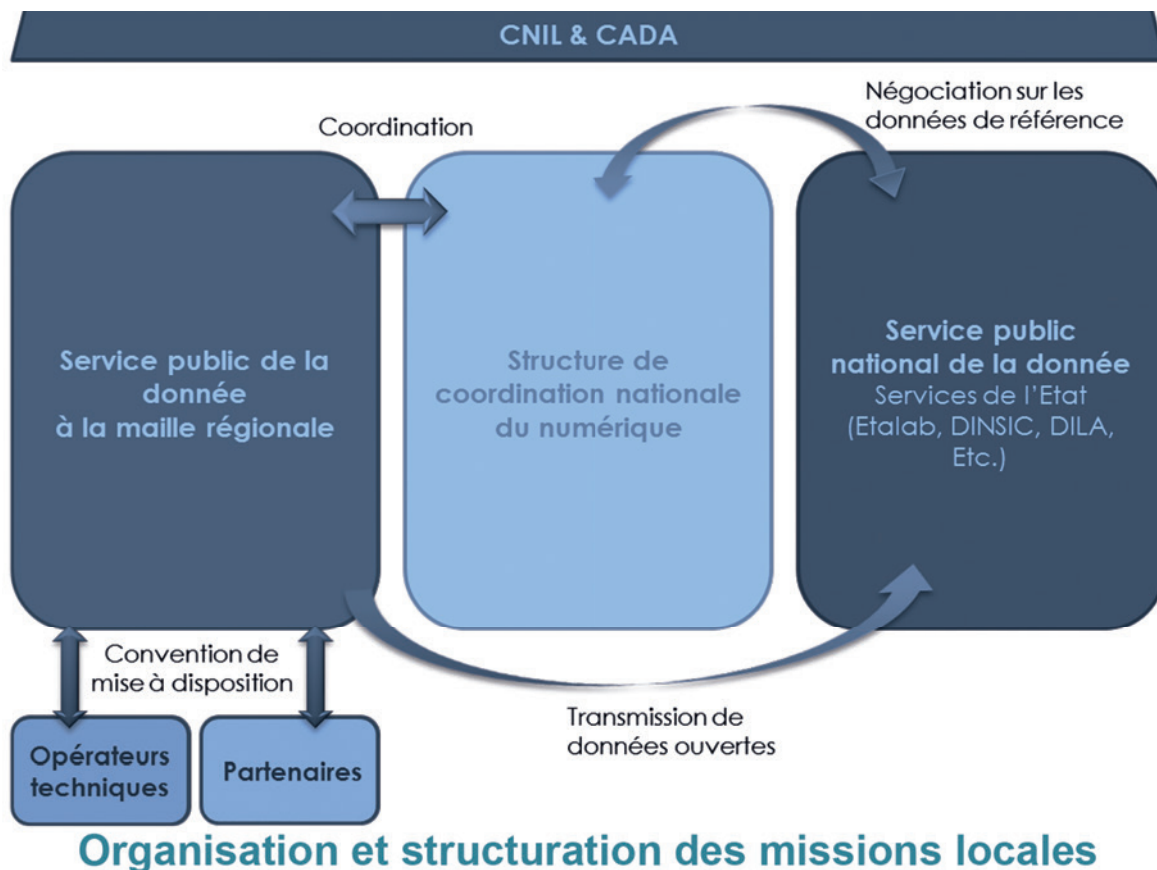
### Quatrième mission : la coopération avec les autres collectivités et avec le secteur privé

Une coopération qui est menée :

- avec les métropoles et les agglomérations du territoire régional, notamment dans le cadre d'échanges de bonnes pratiques et de savoir-faire et dans l'exploration des domaines d'application des *data sciences*,
- avec les autres structures régionales : une coopération interrégionale entre les 13 nouvelles entités régionales pourra être organisée au sein d'une structure de coordination nationale axée sur le numérique et qui pourrait inclure les métropoles. Celle-ci permettrait le travail entre acteurs à la maille régionale sans intervention de l'État, d'aller au-delà de la question des données pour inclure tous les aspects numériques (télécom et applications),
- avec le niveau national : les structures portant les services publics locaux de la donnée pourront prendre part à l'instance de concertation du futur service national de la donnée (par exemple, à travers la structure de coordination nationale du numérique),
- avec le privé et la recherche, au travers de partenariats avec les entreprises, les laboratoires de recherche et le monde universitaire, ainsi qu'avec les associations (médiation numérique, consommateurs, etc.).

Afin d'assurer la gouvernance de la donnée sur le territoire régional concerné, le service décentralisé se dotera d'une instance de pilotage chargée du déploiement d'une infrastructure de données publiques sur le territoire, de la

(3) La suppression de la clause de compétence générale au niveau de la région impliquera de fait la nécessité pour le montage de certains projets d'obtenir un consensus, que ce soit par la concertation, le dialogue ou la collaboration.



## Organisation et structuration des missions locales

valorisation des données publiques, de la diffusion des nouveaux usages de la donnée et de la définition de nouvelles stratégies d'action.

La structure devra par ailleurs acquérir des compétences dans le domaine de la maîtrise d'ouvrage en matière de données. En effet, la transformation des organisations publiques et la mise en place de nouveaux écosystèmes d'information et de nouvelles architectures supposent une maîtrise à la fois des méthodes et des moyens nécessaires à la mise en œuvre de ces ressources. Pour ce faire, le service public local pourra s'appuyer sur des compétences et des opérateurs extérieurs pour expérimenter, développer de nouveaux projets ayant valeur d'exemple et accompagner les acteurs régionaux dans le développement d'une gouvernance de la donnée. La forme juridique de cette structure sera à géométrie variable pour lui permettre de s'adapter au contexte et au jeu des acteurs locaux.

Les entités suivantes pourraient prendre part à la gouvernance de la structure en charge de cette mission locale de gouvernance de la donnée :

- l'État (à travers ses préfetures et ses services déconcentrés),
- les collectivités territoriales (régions, départements, les établissements publics de coopération intercommunale – EPCI (notamment les métropoles)),
- les opérateurs techniques publics (ceux-ci assurant la représentation des autres collectivités du bloc communal),

- les établissements publics (hôpitaux, universités, services départementaux d'incendie et de secours – SDIS, etc.), qui gèrent souvent une grande quantité de données,
- les opérateurs privés locaux,
- les centres régionaux de l'information géographique (CRIGE) et les infrastructures de données géographiques,
- les citoyens ou leurs représentants.

La gouvernance pourrait par exemple fonctionner avec un ou plusieurs collèges, en fonction du contexte. Ainsi, par exemple, le groupement d'intérêt public e-bourgogne est constitué de 13 collèges disposant chacun d'une voix, à l'exception du collège 1, le collège des membres fondateurs qui disposent chacun d'une voix.

Les opérateurs techniques recouvrent l'ensemble des structures locales dotées d'une capacité en ingénierie de projets et de services numériques. Ils peuvent disposer d'une plateforme de services et de moyens d'assistance et d'accompagnement des collectivités territoriales dans la mise en place d'une gouvernance de la donnée à l'échelle régionale incluant de nombreuses entités publiques et, le cas échéant, privées.

Ils représentent une diversité d'acteurs tant publics que privés :

- les opérateurs publics : il peut s'agir des directions des systèmes d'information (DSI) mutualisées des collectivités, des services informatiques des départements, des

syndicats informatiques de mutualisation, des centres de gestion, des agences techniques départementales, ainsi que de certaines structures particulières pouvant parfois revêtir une forme associative,

- les opérateurs privés, parmi lesquels figurent notamment des éditeurs engagés dans la constitution d'infrastructures de partage de données,
- les opérateurs publics-privés rassemblés au sein d'associations de droit privé ou de sociétés coopératives d'intérêt collectif (SCIC), comme se prépare à le devenir l'association gestionnaire de la plateforme Ozwillio.

Assurant la mise en œuvre opérationnelle des projets portés par les missions locales de service public de la donnée, ces structures seront en mesure de jouer un rôle en matière de maîtrise d'ouvrage ou d'assistance à la maîtrise d'ouvrage dans le pilotage de la transition numérique, dans la mise en réseau des acteurs et dans le développement d'une infrastructure de données publiques.