

# Quelle politique de la donnée pour quelles politiques publiques ?

## Ce qu'il faut retenir du Grand Débat de la Donnée du 4 avril 2018



Un événement organisé par

**Syntec**  
NUMÉRIQUE

Sous le haut patronage



En association avec



**CNIL.**  
COMMISSION NATIONALE  
INFORMATIQUE & LIBERTÉS

**etalab** openlab

**FRANCE**   
**URBAINE**  
MÉTROPOLIS, AGGLOS ET GRANDES VILLES

### « La valorisation des données publiques est au cœur des débats (Godefroy de Bentzmann) »

La publication du rapport Villani a placé le sujet de la valorisation des données au cœur des débats, a souligné Godefroy de Bentzmann en introduction du Grand Débat de la Donnée du 4 avril 2018. Gouvernance, impact sur l'action publique, ouverture de nouveaux champs économiques et repositionnement du secteur privé : tels sont les principaux enjeux de la valorisation des données publiques abordés au cours de ce débat initié par Syntec Numérique en association avec la CNIL, l'ANSSI, Etalab ou encore France Urbaine.

### « La circulation des données publiques rend possible la création de valeur et l'innovation » (Simon Chignard)

*Autrefois pensées comme support de la relation entre le public et l'administration, les données publiques deviennent source de création de valeur et d'innovation économique et sociale : tel a été le principal enseignement de la masterclass de Simon Chignard, co-auteur de Datanomics.*

**Une donnée publique est une donnée produite dans le cadre d'une mission de service public**, soit par un acteur public, soit par un acteur privé dans le contexte d'une délégation de service public.

Comme pour toute donnée, **la valeur de la donnée publique est liée à sa circulation, et elle a trois facettes :**

- **la donnée matière première**, qui se vend et s'achète sur une place de marché : l'enjeu n'est plus de monétiser directement la donnée comme matière première, et le temps où l'Etat vendait de grands référentiels de données est révolu. Pour autant, la valeur de la donnée comme matière première **reprend tout son sens dans le contexte de l'essor des technologies d'intelligence artificielle** basées sur l'apprentissage automatique (machine learning, deep learning). En effet, les données publiques sont une source pour les algorithmes apprenants ;
- **la donnée levier d'innovation**, qui est utilisée pour décider, produire, ou agir autrement : **de nombreux innovateurs s'emparent des données publiques** pour construire des services et répondre à des besoins qui n'étaient pas complètement satisfaits, comme le créateur d'Handisco, canne 2.0 pour les malvoyants qui intègre la localisation des arrêts de bus et les passages des bus ;
- **la donnée actif stratégique**, qui donne à celui qui la contrôle une place stratégique dans un écosystème : dans le monde numérique ouvert, **ce n'est pas ce que l'Etat désigne comme référence qui fait référence mais ce qui est utilisé communément partout**. Nous sommes dans l'ère des standards de fait.

En France, dans le contexte de la mise en place à partir d'octobre 2018 de l'open data par défaut, en application de la loi pour une République numérique, le service public de la donnée initié par l'Etat vise à mettre à disposition des données de références avec des garanties de fiabilité, de qualité et de taux d'engagement. L'objectif est de mieux réutiliser les bases de données publiques sans porter atteinte à la vie privée des individus.

### « L'Etat peut penser et agir comme une plateforme » (Henri Verdier)

*La puissance publique peut penser et agir comme une plateforme, et créer de la valeur d'intérêt général : c'est la conviction profonde exprimée par Henri Verdier, directeur interministériel du numérique et du système d'information et de communication de l'Etat (Dinsic), dans sa keynote.*

Pour Henri Verdier, nous sommes dans un monde où **les grandes plateformes deviennent progressivement les bases arrière de l'économie**, ce qui rend quasiment impossible d'innover réellement sans leur accord. Dans ce contexte, la puissance publique a le devoir de comprendre comment ces stratégies de plateforme créent réellement de la valeur mais aussi de garantir la souveraineté de l'Etat et la liberté d'innover de chacun. C'est la raison d'être de

**l'Etat plateforme**, qui a pour objectif de créer de la valeur pour défendre l'intérêt général et de permettre aux entreprises d'innover dans des systèmes économiques où nous avons encore un droit de regard collectif.

L'Etat comporte trois dimensions :

- **la première, c'est l'ambition d'open data**, qui concerne un quart des données publiques environ, non protégées par des secrets légaux (secret des affaires, secret statistique, secret défense...). Elles sont diffusées gratuitement, dans des formats ouverts, avec une liberté totale de réutilisation. Le service public de la donnée doit permettre d'en assurer la fraîcheur, la qualité et la disponibilité 24 heures sur 24. Dans un second temps, l'objectif est de **faire émerger des communs** produits par l'alliance entre l'Etat, les producteurs et les utilisateurs de données, à l'image du partenariat en cours sur la base nationale adresses avec l'Institut national de l'information géographique et forestière (IGN), Open Street Map et La Poste ;
- **la deuxième, c'est la stratégie de plateformes des systèmes d'information de l'Etat**, qui s'appuie notamment sur la mise au point d'API de grande ampleur, pour les entreprises comme les particuliers. Cette stratégie implique également de savoir gérer la **confiance** (certificats de confiance) et l'**identité** (mise en place du *single sign on* avec France Connect, et ouverture progressive de ce service au secteur privé par cercles concentriques). Enfin, elle repose sur des standards ouverts, avec des codes sources publiés.
- **la troisième, c'est que l'Etat apprenne à innover avec la plateforme et à créer de la valeur par des méthodes agiles.**

### « Transformation de l'action publique par la donnée : vers une coopération entre acteurs publics et privés »

*Avec : Guillaume Blot, chief digital and innovation officer, Sopra Steria Banking Software, Benjamin Jean, président d'Inno<sup>3</sup>, Camille Putois, fondatrice de K, et Henri Verdier, directeur interministériel du numérique et du système d'information et de communication de l'État (DINSIC).*

**La révolution, ce n'est pas l'ouverture des données publiques mais les collaborations qui naissent entre acteurs publics, privés et associatifs pour renforcer cet actif et imaginer de nouveaux services : tel a été le fil rouge de la première table ronde du Grand Débat de la Donnée.**

Le secteur privé n'a pas attendu l'Etat pour mener des politiques d'open data, et ce qui change, c'est l'invention de **nouvelles formes de collaboration entre acteurs publics et privés**, pour Benjamin Jean, président d'Inno<sup>3</sup>. Ainsi, dans les transports, l'ensemble des acteurs de la mobilité se sont réunis pour fournir une base de données mutualisée qui puisse faire référence. Différents modèles de circulation des données émergent : tandis que certaines entreprises ouvrent l'ensemble de leurs données dans une démarche collaborative, d'autres participent à des projets initiés par les acteurs publics, dans une logique de **coopétition public-privé**.

De son côté, Henri Verdier, directeur interministériel du numérique et du système d'information et de communication de l'État, distingue **différents cas de figures appelant différentes réponses** : les données provenant du secteur public ou privé qui peuvent être ouvertes sans danger, et les données provenant du secteur public ou privé qui nécessitent la protection d'un secret légal. La **notion de données d'intérêt général**, qui émerge aujourd'hui, implique que certaines données issues du secteur privé puissent être utilisées par la puissance publique, par exemple afin d'organiser les transports dans une ville. Mais **la révolution de l'intelligence artificielle va bousculer cet état de fait** : il va falloir mener plus de recherches sur des données nécessitant la protection d'un secret légal, y compris celles issues du secteur privé. Cela ne signifie pas que cela se fera à titre gratuit.

La mise en circulation des données va permettre de **réécrire complètement les parcours usagers autour de services publics numériques**, et les secteurs économiques qui sont également dans une logique de plateforme auront un rôle majeur à jouer dans cette évolution :

c'est la conviction de Guillaume Blot, chief digital and innovation officer de Sopra Steria Banking Software. Des API permettront de réinventer les interactions entre public et privé en simplifiant la vie des usagers, par exemple pour la fourniture de justificatifs de domicile ou de relevés fiscaux.

Les API devront également à terme permettre d'aller vers une **coproduction des démarches administratives**, selon Camille Putois, fondatrice de K. Les systèmes d'information qui gèrent les procédures pourraient s'ouvrir pour que des partenaires extérieurs proposent leur propre version de la démarche administrative et intègrent une partie du système d'information concerné dans le service qu'elles proposent, par exemple pour la déclaration de revenus ou les titres de séjour. Des garde-fous sont toutefois nécessaires : gratuité de la démarche pour l'utilisateur, respect de la vie privée, accès de l'administration à l'ensemble des données d'activité liées à cette démarche. Cette **diversification des interfaces** est importante **pour l'inclusion** de tous les citoyens, a rappelé Henri Verdier, et il faut des **données de référence** sur lesquelles tout le monde se synchronise et qui deviendront des standards.

Une logique de **communs**, en tant que **ressource partagée, avec des règles de gouvernance permettant de la pérenniser**, est ainsi en train d'émerger dans tous les secteurs, a ajouté Benjamin Jean, notamment dans la mobilité, ou dans le secteur spatial. Reste à résoudre la question de la valorisation comptable de cette ressource partagée. De **nouveaux business models sont à imaginer**, a conclu Henri Verdier.

### **Circulation des données : les services publics réinventés**

*Avec Henri Isaac, président de Renaissance Numérique, Jean-Marc Lazard, président d'OpenDataSoft, Anne-Claude Poinso, Responsable développement Collectivités Microsoft France, et Sarah Prot, Responsable des affaires publiques d'Airbnb en France.*

**Les conditions de mise en place d'une politique d'open data par les collectivités et l'impact des données produites par certains acteurs privés dans la modification des politiques publiques : tels ont été les deux principaux sujets de la deuxième table ronde du Grand Débat de la Donnée.**

#### *Partir des usages pour définir une stratégie d'open data*

La publication des données en tant que telle n'est pas prioritaire pour nombreuses collectivités, car elle nécessite de mobiliser des ressources humaines et financières alors que le retour sur investissement n'est pas immédiat : c'est ce qu'a souligné Anne-Claude Poinso, responsable développement collectivités chez Microsoft France. Les collectivités doivent donc prendre conscience que les données sont leur patrimoine informationnel, et **partir des attentes des citoyens** en matière d'information et d'accès aux services pour élaborer leur politique d'ouverture des données. Il leur faut également évaluer de quelles données elles disposent, et identifier parmi ces données celles qui sont qualitatives, réutilisables et peuvent être mises à jour.

Partir des usages est également la démarche conseillée par Jean-Marc Lazard, président d'OpenDataSoft, car la donnée est anxiogène. Cette approche se justifie aussi par le fait qu'**il n'y a pas un modèle d'organisation unique** : chaque territoire a sa problématique propre. Le croisement des données entre collectivités pourra d'ailleurs permettre d'imaginer des services innovants, a fait remarquer Anne-Claude Poinso.

Sur le plan de la gouvernance, au-delà de la nomination d'un chief data officer, pour Henri Isaac, président de Renaissance Numérique, ce qui compte c'est de mettre en place des **processus qui garantissent la qualité des données**, depuis leur entrée dans le système : c'est donc l'ensemble des acteurs qui gèrent la donnée qui doivent être acculturés. De son côté, Jean-Marc Lazard souligne que l'ouverture des données publiques fait entrer la transversalité dans un système à structure pyramidale : c'est un **changement culturel à accompagner**.

Enfin, la circulation des données ne pourra se faire que si elles sont standardisées, a souligné Henri Isaac, tandis que Jean-Marc Lazard expliquait que cette standardisation ne se fait que par rapport à l'usage qu'elle permet : c'est la **valeur d'usage des données** qui va faire faire des gestes rétroactifs à toute la chaîne de valeur.

#### *Partage des données des acteurs privés et émergence de nouveaux services publics*

Au-delà de l'ouverture de leurs propres données, l'accès à certaines données produites par des acteurs privés est une demande des acteurs publics. Ainsi, AirBnB est sollicité par de nombreuses communes pour avoir des informations leur permettant de mieux mesurer l'activité touristique sur leur territoire : en réponse à cette demande, AirBnB a développé un outil, Dataville, qui leur donne accès à certaines données clés, comme le nombre d'hébergements proposés sur leur territoire, le nombre de voyageurs, etc., a expliqué Sarah Prot, responsable affaires publiques d'AirBnB en France. L'entreprise est par ailleurs engagée dans une démarche globale de valorisation de ses données à travers des partenariats avec les collectivités territoriales, notamment les régions et départements.

Cette collaboration entre acteurs privés et publics est indispensable pour diminuer les externalités négatives de certains services numériques sur la collectivité, a souligné Henri Isaac. Par exemple, Waze ne réduit pas les embouteillages mais les déplace dans des zones résidentielles qui ne sont pas dimensionnées pour les recevoir, selon une étude de l'université de Berkeley. Le service pourrait être amélioré en y intégrant les données publiques sur les travaux en cours : c'est la **combinaison de jeux de données privés et publics** qui permet de concilier intérêt public et intérêt privé.

On peut même aller plus loin en matière de circulation des données, selon Jean-Marc Lazard. A un premier niveau, Waze fournit des données en temps réel à la collectivité, qui, de son côté, lui transmet les informations sur les travaux en cours, la localisation des camions poubelles, etc, lui permettant d'affiner ses propositions d'itinéraire. Mais à un deuxième niveau, le stockage des données permet à la collectivité de les analyser et de planifier différemment sa politique de voirie. Ainsi, **les données de l'entreprise vont influencer sur les services publics**.

Pour conclure, Henri Isaac a fait remarquer que **les entreprises sont prêtes au partage des données si elles peuvent les réutiliser**.

#### **« Maturité et pragmatisme » (Godefroy de Bentzmann)**

En conclusion, Godefroy de Bentzmann a souligné que les débats avaient montré une **montée en maturité progressive**, avec une **approche pragmatique** en matière de gouvernance des données publiques. La **circulation des données** ne se fait pas dans un seul sens, et elle a **des conséquences majeures sur la relation entre entreprises et organismes publics**.