

Le service public des données de référence : enjeux de gouvernance

Par Mathilde HOANG et Antonin GARRONE
DINUM, Etalab

À l'heure où la Commission européenne dresse la liste des ensembles de données à haute valeur qui devront faire l'objet d'une diffusion au grand public, il apparaît nécessaire de mettre en question les enjeux de gouvernance d'un tel dispositif. Le service public des données de référence, créé dans le cadre de la loi pour une République numérique (2016), peut apporter des enseignements sur les critères de succès en matière de gouvernance pour ces deux dispositifs. Dans le cas du modèle français, la création d'un service public par voie légale, ainsi que la formalisation d'un ensemble de règles organisationnelles et techniques qui encadrent la diffusion et la réutilisation des données ont entraîné la création d'une gouvernance multipartite ainsi que la mise en place d'une infrastructure de données fiable et performante. La mesure de l'effet du dispositif, ainsi que la prise en compte des évolutions de la politique de la donnée au niveau national et international devront guider l'évolution du service public.

Dans le cadre de l'étude d'impact de la directive n°2019/1024/UE du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public (dite PSI III), la Commission européenne affirme que l'impact de l'ouverture des données est directement influencé par des facteurs tels que les coûts, la qualité des données, leur documentation et leurs modalités d'accès¹. Dans cette perspective, elle considère qu'il est nécessaire de prioriser les efforts portés sur ces facteurs pour des ensembles de données de forte valeur, relevant de thématiques particulières². Ces ensembles de données devront être mis à disposition dans des formats lisibles par machine, en recourant à des interfaces de programmation appropriées. Alors que la Commission travaille de concert avec ses partenaires pour déterminer la liste des ensembles de « données de haute valeur », il apparaît pertinent de s'attarder sur le modèle français du service public des données de référence, dont les enjeux de gouvernance font écho à ceux du dispositif européen³. Le service public des données de référence vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social. En 2016, une première liste de neuf jeux de données de référence a été établie : quatre d'entre eux permettent d'identifier des entreprises, des associations, des

¹ Commission européenne (2019), "Impact assessment study on the list of high value datasets to be made available by the Member States under the Open Data Directive".

² À ce stade, les thématiques pressenties sont les catégories géospatiales, observation de la terre et environnement, météorologiques, statistiques, entreprises et propriété d'entreprises, et mobilité (source : directive PSI III).

³ La gouvernance des données est définie par la Commission européenne comme un « ensemble de règles et de moyens applicables à l'utilisation des données, s'appuyant par exemple sur des mécanismes de partage, des accords et des normes techniques » (source : Commission européenne (2020), « Stratégie européenne pour les données », COM, 19 février).

administrations, des métiers et emplois, et les cinq autres constituent un corpus relatif aux données géographiques⁴.

Si le parallèle entre les dispositifs européens et français doit se faire avec nuance (le périmètre des données, la liste des partenaires ou encore les critères de qualité attendus ne sont pas toujours comparables), il reste utile de s'interroger sur les facteurs en matière de gouvernance qui permettent d'assurer le succès de tels dispositifs⁵.

À cet égard, la création du service public des données de référence, accompagné d'un ensemble de règles et de moyens applicables à l'utilisation des données, a entraîné l'émergence d'une gouvernance multipartite et d'une infrastructure de la donnée indispensable à la diffusion et la réutilisation des données.

INSTAURER PAR VOIE LÉGALE UN SERVICE PUBLIC DES DONNÉES DE RÉFÉRENCE

Si la diffusion et la réutilisation des données publiques sont le fruit d'une évolution progressive du cadre légal français et européen⁶, c'est la loi pour une République numérique du 7 octobre 2016 et, quelques temps auparavant, la loi du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public qui ont entraîné un changement de paradigme significatif en consacrant de nouveaux principes en matière d'ouverture des données publiques : instauration du principe d'ouverture des données par défaut des données publiques, généralisation de la gratuité des données publiques ou encore le principe de réutilisation libre des données⁷.

Au-delà de ce cadre général favorable à l'ouverture et la réutilisation des données publiques, le législateur a considéré que certaines données publiques occupaient une place spécifique du fait de leur potentiel d'usage pour l'économie, l'action publique et la société. Dès lors, ces données doivent être régies par un régime particulier, afin de fiabiliser les modalités d'accès à ces données⁸.

Au-delà de la charge symbolique que représente la création d'une nouvelle mission de service public, la constitution juridique du service public des données de référence en tant que mission d'intérêt général entraîne l'élaboration et l'application d'un cadre propre⁹. À ce titre, l'article 14 de la loi pour une République numérique consacre la création d'une nouvelle mission de service public confiée à l'État : le service public des données de

⁴ La liste complète des données de référence est la suivante : le répertoire des entreprises et de leurs établissements, le répertoire national des associations, la base de l'organisation administrative, le code officiel géographique, le plan cadastral informatisé, le registre parcellaire graphique, le référentiel à grande échelle, la base adresse nationale (BAN), le répertoire opérationnel des métiers et des emplois (source : Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence).

⁵ Dans cette perspective, il convient de souligner que l'enjeu de cet article n'est pas de dresser les bénéfices attendus du service public de la donnée ou d'établir son bilan, mais de tirer des enseignements quant aux critères de succès de la mise en œuvre de ce dernier.

⁶ La loi dite CADA (Commission d'accès aux documents administratifs) du 17 juillet 1978 a affirmé la liberté d'accès aux documents administratifs et a notamment été modifiée en 2005 pour introduire, en application de la directive de 2003, le principe de réutilisation des données publiques.

⁷ La réutilisation libre des données signifie que celle-ci est gratuite, qu'elle peut viser une autre finalité que le but initial de production des données et qu'elle peut être réalisée par tout acteur (public comme privé).

⁸ Assemblée nationale (2015), « Étude d'impact, projet de loi pour une République numérique », 9 décembre.

⁹ Cluzel-Métayer (2018), « La construction d'un service public de la donnée », *Revue française d'administration publique*, 167(3), pp. 491-500.

référence, dont la mission est de mettre à disposition, en vue de faciliter leur réutilisation, les données de référence¹⁰. La notion de donnée de référence reconnaît le statut particulier de ces bases de données répondant à trois critères : elles servent à identifier ou nommer des produits, des services, des lieux et des personnes ; elles sont utilisées fréquemment par des acteurs publics ou privés autres que l'administration qui les détient ; et la qualité de leur mise à disposition est critique pour ces utilisations.

L'article précité représente la première pierre de l'édifice juridique du service public des données de référence. La formalisation du périmètre et de l'objet du service public, ainsi que des critères de définition des données de référence est indispensable pour établir par la suite les conditions de diffusion et de réutilisation des données de référence. La loi s'accompagne de fait d'un ensemble de mesures de nature réglementaire, qui décrivent les modalités de mise en œuvre du service public. C'est à partir de ces règles qu'ont été établies les modalités de coordination entre acteurs ainsi que les critères techniques de diffusion des données.

IDENTIFIER ET COORDONNER LES RÔLES DES PARTIES PRENANTES

Les données de référence sont produites par une variété d'acteurs. Au regard de la diversité des parties prenantes, le législateur a institué par voie réglementaire un schéma d'organisation qui vise une gouvernance efficace du service public¹¹. Les producteurs des données de référence sont chargés de la production, de la documentation, de la mise à jour des données ainsi que de l'identification de l'acteur responsable de la diffusion des données. Ce dernier (qui peut être le producteur lui-même) les diffuse dans le respect des critères de performance et de disponibilité fixés par arrêté. En sus de ces acteurs aux missions et compétences plurielles, le département Etalab¹², de la direction interministérielle du numérique de l'État (DINUM), concourt au pilotage du service public des données de référence en coordonnant la mise à disposition des données. Il veille notamment à la fiabilité, la disponibilité, la sécurité et la performance des services, et participe à l'amélioration de la qualité des données à travers des échanges avec les réutilisateurs et producteurs de données. Par ailleurs, il favorise l'émergence de services innovants issus de la réutilisation des données et cherche à faire évoluer la liste des données de référence au sein du service public des données de référence.

Si les engagements relatifs à la diffusion des données de référence relèvent de la responsabilité d'acteurs publics, les utilisateurs du service public des données de référence sont aussi partie prenante du dispositif. Puisque le succès de ce service public repose sur la réutilisation effective des données, ce sont leurs usagers potentiels qui ont été sollicités pour définir les contours opérationnels du service public des données de référence. La mission Etalab a ainsi mené une consultation, à laquelle 160 répondants ont contribué, afin d'établir une première liste de données de référence ainsi que les critères de qualité et de disponibilité attendus¹³. Au-delà de la contribution à la construction du service public,

¹⁰ Loi n°2016-1321 du 7 octobre 2016 pour une République numérique, article 14.

¹¹ Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence.

¹² Etalab est un département de la direction interministérielle du numérique de l'État (DINUM) qui est une direction sous l'autorité de la ministre de la Transformation et de la Fonction publiques. Le département coordonne et promeut l'action de l'État et des organismes placés sous sa tutelle en matière d'inventaire, de gouvernance, de production, de circulation, d'exploitation et d'ouverture des données, et notamment des codes sources.

¹³ Consultation en ligne entre le 29 septembre et le 20 octobre 2016 dont les résultats sont disponibles sur la plateforme [data.gouv.fr](https://www.data.gouv.fr), <https://www.data.gouv.fr/fr/datasets/reponses-a-la-consultation-sur-la-mise-en-oeuvre-du-service-public-de-la-donnee/>

les utilisateurs des données participent à l'amélioration de la qualité de celles-ci en signalant des erreurs à propos des données ou en proposant des améliorations techniques ou organisationnelles. Ils contribuent enfin à l'élaboration de nouveaux services innovants à partir des réutilisations qu'ils réalisent des données.

L'identification de chaque partie prenante ainsi que ses rôles et responsabilités assoit une gouvernance multipartite dont les contours sont clairs, mais restent évolutifs. L'implication des acteurs dans la co-construction et l'animation du service public est un facteur indispensable au succès du dispositif. C'est en s'appuyant sur cette gouvernance¹⁴ que des engagements techniques peuvent être identifiés afin d'assurer la diffusion, la disponibilité, la performance et la qualité du service public des données de référence¹⁵.

CONSTRUIRE UNE INFRASTRUCTURE DE LA DONNÉE

Les bénéfices de l'*open data*, et *a fortiori* du service public de la donnée, reposent sur la capacité des utilisateurs à accéder et réutiliser facilement les données publiques. Pour cela, un ensemble de règles techniques a été établi pour favoriser l'émergence d'une infrastructure de la donnée capable de garantir la fourniture de données à un niveau de qualité de service industriel¹⁶.

La disponibilité des données de référence constitue le premier critère de fiabilité de l'infrastructure. Si le service est interrompu ou dégradé, la confiance des utilisateurs diminue, ces derniers n'étant pas en capacité de concevoir des services fiables à partir des données mises à disposition. Il est ainsi demandé aux producteurs de données de respecter un taux de disponibilité du service et un niveau de performance excellent¹⁷. Le respect de ces critères peut être facilité par la mise en place d'une infrastructure commune de diffusion des données. En cela, la plateforme nationale de données publiques « data.gouv.fr »¹⁸, qui permet la mise à disposition des données publiques dans des modalités simples, fait office de service à destination des producteurs de données, qui ne sont ainsi pas tenus de mettre en œuvre une infrastructure dédiée.

À la disponibilité s'ajoute l'enjeu de découvrabilité des données¹⁹, facteur clé pour leur réutilisation. Le référencement des données de référence sur le portail data.gouv.fr favorise l'identification par les réutilisateurs des données à fort potentiel d'impact. L'obligation pour les producteurs de référencer les informations clés relatives aux données

¹⁴ Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence.

¹⁵ Arrêté du 14 juin 2017 relatif aux règles techniques et d'organisation de mise à disposition des données de référence prévues à l'article L.321-4 du Code des relations entre le public et l'administration.

¹⁶ Arrêté du 14 juin 2017 relatif aux règles techniques et d'organisation de mise à disposition des données de référence prévues à l'article L.321-4 du Code des relations entre le public et l'administration.

¹⁷ L'administration chargée de la mise à disposition de données de référence garantit la disponibilité des données en téléchargement 99 % du temps mensuel, apprécié au terme de chaque mois. Elle s'efforce de garantir la disponibilité des données par l'intermédiaire d'une interface de programmation 99,5 % du temps mensuel, apprécié au terme de chaque mois.

¹⁸ La plateforme data.gouv.fr est désignée par le législateur comme le portail unique interministériel (source : Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence).

¹⁹ Le terme de découvrabilité est souvent utilisé pour qualifier le degré auquel une donnée peut être trouvée lors d'une recherche, ou la capacité des utilisatrices et utilisateurs à découvrir une donnée dont ils ignoraient l'existence.

(métadonnées) favorise également la découverte des données ainsi que leur prise en main par les usagers du service public²⁰.

Des modalités de mise à disposition des données de référence sont également établies afin d'assurer un accès adapté aux données de référence. Ces dernières doivent être disponibles en téléchargement dans leur intégralité, être facilement réutilisables et, le cas échéant, par l'intermédiaire d'interfaces de programmation (API). À ce sujet, des API (API SIRENE, API RNA, API Geo) ont été conçues par le département Etalab afin de faciliter l'accès à d'importants volumes de données.

Pour autant, la diffusion des données publiques, même accompagnée d'un cadre légal et d'une gouvernance formelle, n'entraîne pas directement leur réutilisation. C'est pour répondre à ce constat que des exigences de qualité ont été établies²¹.

Plusieurs éléments permettent d'évaluer le niveau de qualité d'un jeu de données. Il peut s'agir d'éléments sur les données elles-mêmes et sur leur structure (le format, structure et contenu du fichier, etc.), des éléments attestant du potentiel de réutilisation et de croisement des données (le respect de standards et schémas déjà établis, la présence de données pivots pour lier les données à un référentiel, etc.) ou encore des éléments qui apportent des informations sur les données (documentation, gestion des versions et des mises à jour, etc.).

Dans le cadre du service public des données de référence, afin d'assurer un niveau de qualité satisfaisant, les producteurs sont notamment tenus de renseigner les métadonnées relatives aux données de référence, d'indiquer la périodicité de la mise à jour et d'apporter des informations sur le processus de création des données.

La réutilisabilité des données est également dépendante de la capacité des parties prenantes à échanger. Afin d'assurer la continuité des échanges, les producteurs des données sont tenus de répondre aux questions des utilisateurs dans un délai maximal d'un mois et d'informer le public en amont d'une modification dans le processus de diffusion ou dans les données elles-mêmes.

L'infrastructure des données peut faciliter les interactions, que ce soit en proposant des espaces d'échanges (forums, espaces de commentaires) ou en encourageant le référencement des réutilisations. Par ailleurs, la connaissance des usages favorise les échanges entre réutilisateurs et producteurs de données, et peut insuffler des collaborations. Les producteurs sont encouragés à répondre aux réutilisateurs identifiés dont ils saisissent mieux le besoin. Les réutilisations permettent ainsi d'améliorer la qualité d'un jeu de données en facilitant le partage d'informations sur les manques et les anomalies dans les données.

Etalab est directement partie prenante dans cette dynamique d'échanges et d'animation de la communauté. Au-delà d'assurer la mise en relation entre producteurs et réutilisateurs de données, le département doit également contribuer à l'émergence de services innovants issus des réutilisations des données²². À ce titre, il met à disposition des

²⁰ Les informations à référencer sont listées dans l'arrêté du 14 juin 2017 relatif aux règles techniques et d'organisation de mise à disposition des données de référence prévues à l'article L.321-4 du Code des relations entre le public et l'administration.

²¹ Arrêté du 14 juin 2017 relatif aux règles techniques et d'organisation de mise à disposition des données de référence prévues à l'article L.321-4 du Code des relations entre le public et l'administration.

²² Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence.

services qui facilitent la prise en main des données, tels que des outils de géocodage²³ (pour la base adresse nationale et la base SIRENE des entreprises) ou encore une suite d'outils autour du plan cadastral qui enrichissent les données brutes et facilitent le téléchargement de certaines parties du cadastre.

MESURER LES IMPACTS ET FAIRE ÉVOLUER LE DISPOSITIF

La mise en place d'un cadre favorable à l'appropriation des données du service public des données de référence, que ce soit en termes organisationnel et technique, doit se faire de manière conjointe avec la mesure des freins aux usages et à l'impact de ceux-ci.

La mesure des usages des données ouvertes, et en particulier des données de référence, est essentielle pour estimer l'impact de l'ouverture et de la diffusion des données. L'évaluation des impacts des données ouvertes est un sujet complexe auquel se confronte l'ensemble des parties prenantes. Ceci peut s'expliquer par plusieurs raisons. Premièrement, le principe fondateur de l'*open data* est d'adopter un positionnement agnostique selon lequel il ne faut ni préjuger ni contrôler les usages futurs qui seront faits des données mises à disposition. Ensuite, la tâche d'évaluation exhaustive des effets directs et indirects de l'ensemble des retombées (sociales, politiques, économiques, etc.) représente un effort considérable et demeure relativement peu mise en œuvre.

La mesure plus systématique des usages et des effets peut être considérée comme un facteur de succès et demeure insuffisamment mise en œuvre. Il s'agit non seulement de mesurer quantitativement l'intérêt pour ces données²⁴, mais également de suivre qualitativement les usages. Il faut s'interroger sur les services et l'usage des services²⁵ qui s'appuient sur l'infrastructure de la donnée, sans pour autant risquer de restreindre les usages²⁶.

L'un des trois fondements d'un service public est son adaptabilité, c'est-à-dire sa capacité à faire face aux évolutions de la société ainsi qu'aux évolutions techniques. Après cinq années de mise en œuvre, le service public des données de référence s'intègre désormais dans un nouvel écosystème de la donnée. La crise sanitaire a révélé le besoin critique de produire, partager et exploiter des données publiques de qualité, bien au-delà du périmètre du service public des données de référence. À cet égard, la circulaire du Premier ministre relative à la politique publique de la donnée, des algorithmes et des codes sources a réaffirmé l'ambition gouvernementale de produire des données de qualité et de les faire circuler auprès de l'ensemble des acteurs habilités à en connaître²⁷. Une nouvelle gouvernance de la donnée a également été instaurée avec la nomination pour chaque ministère d'un/e administrateur/trice ministériel/le des algorithmes et des codes sources chargé/e de mettre en œuvre, pour son périmètre, un ensemble d'engagements en matière de politique de la donnée. Au niveau européen, la mise en œuvre de la directive PSI III, avec en son cœur l'identification des ensembles de données à haute valeur,

²³ Le géocodage consiste à affecter des coordonnées géographiques (longitude/latitude) à une adresse postale.

²⁴ Certains indicateurs comme le nombre de visites, de téléchargements ou encore d'appels aux API peuvent constituer un faisceau d'informations.

²⁵ En effet, il s'agit de mesurer le nombre de réutilisations et leur qualité, mais aussi leur niveau d'utilisation : s'agit-il d'une application utilisée par des millions de personnes ou d'une visualisation de données ponctuelle par exemple.

²⁶ L'usage de licences de réutilisation trop restrictives ou encore l'obligation de s'authentifier pour télécharger les données peuvent par exemple faciliter le suivi des usages, mais peuvent aussi représenter des freins considérables à la réutilisation.

²⁷ Circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources.

ainsi que les réflexions autour des espaces communs de données européens, replace le sujet au niveau international. Elles soulèvent notamment des interrogations relatives à l'interopérabilité des données, à la mise en œuvre d'une gouvernance internationale ou encore à l'émergence d'infrastructures européennes des données.

À l'aune de cette nouvelle dynamique, l'évolution du service public des données de référence, de sa gouvernance, de sa liste de données de référence et de ses critères techniques se pose. L'objectif du service public restera quant à lui le même : créer une infrastructure des données de qualité afin d'améliorer l'action publique, de favoriser la création de nouveaux services innovants et de renforcer la souveraineté numérique nationale.

RÉFÉRENCES BIBLIOGRAPHIQUES

ASSEMBLÉE NATIONALE (2015), « Étude d'impact, projet de loi pour une République numérique », 9 décembre.

Arrêté du 14 juin 2017 relatif aux règles techniques et d'organisation de mise à disposition des données de référence prévues à l'article L.321-4 du Code des relations entre le public et l'administration.

Circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources.

CLUZEL-MÉTAYER (2018), « La construction d'un service public de la donnée », *Revue française d'administration publique*, 167(3), pp. 491-500.

COMMISSION EUROPÉENNE (2019), "Impact assessment study on the list of high value datasets to be made available by the Member States under the Open Data Directive".

COMMISSION EUROPÉENNE (2020), « Stratégie européenne pour les données », COM(2020), 19 février.

Décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence.

Directive n°2019/1024/UE du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur.

Loi n°2016-1321 du 7 octobre 2016 pour une République numérique, article 14.