Le 5 juin 2012

Avis n° 12 du Conseil national du numérique relatif à l'ouverture des données publiques (« Open data »)

Mis en place il y a 34 ans, le droit dont disposent les citoyens d'accéder à la plupart des documents publics répond en premier lieu à un souci de transparence de l'administration. Les évolutions technologiques de ces quinze dernières années devraient permettre, en théorie, une mise à disposition du plus grand nombre, gratuite, de toutes les données publiques. En conséquence, elles ont considérablement accru dans le même temps les attentes, voire les exigences de la société civile. En particulier, celle de pouvoir diffuser et réutiliser ces informations.

Il s'agit d'abord de donner une plus grande lisibilité à l'action publique, et notamment de permettre d'ancrer les débats sur des informations précises. L'ouverture des données rend par exemple possible les exercices de « *benchmark* » qui, en mettant en évidence les bonnes pratiques, dysfonctionnements et disparités, peuvent conduire à une allocation plus efficace et plus juste des ressources publiques. Elle est une **condition de la démocratie participative et une source d'amélioration de l'efficacité des administrations.**

L' « *open data* » est aussi un premier pas vers la construction d'une administration décloisonnée, plus réactive, plus efficace et plus proche des usagers. Une meilleure compréhension du fonctionnement du service public peut permettre à ces derniers d'ajuster et formuler plus finement leurs attentes, et surtout d'y participer. Cette dimension participative est particulièrement pertinente au niveau local : citoyens et associations, mais aussi entrepreneurs pourraient développer des services innovants et contribuer à **améliorer la qualité de vie** de la population.

La mise à disposition des innombrables données collectées et produites par les administrations dans l'exercice de leurs missions de service public révèle progressivement leur valeur économique et sociale. Il est encore trop tôt pour en mesurer les retombées, même si une récente étude[1] estime qu'un accès facilité pourrait augmenter de 40 milliards d'euros le marché européen des informations du secteur public, aujourd'hui estimé à 28 milliards d'euros par an. L'impact économique total, direct et indirect, atteindrait alors 140 milliards d'euros par an. L'« *open data* » constitue ainsi un **enjeu de développement de l'économie numérique**. Véritable **facteur de production**, les données publiques permettent de créer de nouvelles applications, et surtout d'enrichir et mettre en valeur les services existants.

Finalement, le secteur public, qui pèse pour plus de 50% dans l'économie française, peut en mettant à disposition ses données jouer un **rôle moteur dans la construction de la société numérique**, tout en modernisant son propre fonctionnement.

C'est dans ce cadre que le Conseil national du numérique a décidé de s'autosaisir du thème de l'ouverture des données publiques (« Open data »).

[1] *Review of recent studies on PSI re-use and related market developments,* Graham Vickery, 2011. Etude commandée par la Commission Européenne en préparation de la révision de la directive concernant la réutilisation des informations du secteur public. http://ec.europa.eu/information_society/policy/psi/index_en.htm

Des constats essentiels se dégagent :

- La mise à disposition des données publiques est un **impératif démocratique**, un **puissant vecteur de modernisation de l'administration** et un **composant important de l'économie numérique**

- Les initiatives de l'Etat, avec notamment la mission Etalab, et des collectivités territoriales sont très encourageantes et marquent **un tournant dans la politique des données publiques**.

- Il faut maintenant faciliter l'**extension** et la **pérennisation** de cette dynamique naissante, et porter l'ambition à un **niveau stratégique**. La réalisation pleine du potentiel offert par la réutilisation des données publiques requiert une **vision politique,** une **adaptation du cadre législatif** et une évolution **des pratiques de l'administration**.

Les recommandations du CNNum suivent les deux axes d'action complémentaires suivants :

1) **Réformer le droit relatif aux données publiques (Propositions 1 a 5)**
2) **Orienter et accompagner la démarche des administrations (Proposition 6 à 9)**

Et, afin de permettre leur mise en action :

3) **Mettre en place une gouvernance des données publiques (Proposition 10 et 11)**

Historique : de la liberté d'accès aux documents administratifs à « l'Open Data »

Accès aux documents administratifs

La loi « CADA » du 17 juillet 1978[2] pose les fondements de l'ouverture des données publiques en donnant aux citoyens le droit d'accéder à la plupart des documents administratifs, définis comme « *les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'un telle mission* ». Autorité administrative indépendante, la commission d'accès aux documents administratifs (CADA) est dans le même temps créée. Elle est chargée de veiller au respect de ce droit et assure notamment la médiation entre les administrations et les demandeurs, avant un éventuel recours contentieux. A l'ère du papier, le droit est relativement peu utilisé.

Réutilisation libre et gratuite des informations publiques

La vision des choses évolue considérablement avec l'apparition d'internet, qui permet une diffusion des informations à coût marginal quasi nul et l'émergence d'une société de l'information. Le discours d'Hourtin prononcé par le Premier Ministre le 25 août 1997 marque la reconnaissance en France du caractère économique essentiel de certaines données publiques et de la nécessité de les diffuser gratuitement.

[2] Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal

Transposant la directive européenne du 17 novembre 2003[3], l'ordonnance du 6 juin 2005[4] vient ainsi compléter le dispositif créé en 1978 en posant les principes d'une réutilisation libre des informations publiques, ces dernières étant définies comme les informations contenues dans les documents administratifs. Des exceptions notables sont toutefois faites à ce droit, en particulier les activités à caractère industriel et commercial. Son application est bien entendue subordonnée au respect de la vie privée, de la propriété intellectuelle etc.

Vers une mise en ligne des données publiques

La troisième phase marque le passage progressif d'une logique de demande à une logique d'offre avec l'installation le 21 février 2011 de la mission Etalab, chargée de créer un portail unique destiné à mettre à disposition libre l'ensemble des informations publiques de l'Etat et de coordonner dans ce but l'action des administrations[5]. Une circulaire du Premier Ministre du 26 mai 2011[6] précise la démarche à suivre pour faciliter la « *réutilisation libre, facile et gratuite des informations publiques* »[7].

Le portail de l'Etat data.gouv.fr est ainsi ouvert le 5 décembre 2011, offrant lors de son lancement plus de 350 000 données provenant des administrations mais aussi des collectivités territoriales volontaires. Un nombre croissant de collectivités lance d'ailleurs leurs propres initiatives ; quelques-unes ayant même été pionnières (Rennes, Saône-et-Loire, Paris..). Cet effort fait de la France un des pays les plus investis dans ce mouvement désormais mondial[8], aux côtés notamment des Etats-Unis (data.gov ouvert en mai 2009) et de la Grande-Bretagne (data.gov.uk en septembre 2009).

Les propositions du Conseil national du numérique

Les travaux communautaires relatifs à l'ouverture des données publiques, le lancement du portail de l'Etat et la multiplication d'initiatives locales montrent partout une volonté politique croissante d'avancer vers plus de transparence, de confiance et de symbiose entre le secteur public et la société civile. Les nombreux projets « open data », en France ou partout dans le monde illustrent concrètement les possibilités offertes par cette tendance.

---

[3] Directive 2003/98/CE du 17 novembre 2009 concernant la réutilisation des informations du secteur public.

[4] Ordonnance n°2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

[5] Décret n°2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques

[6] circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'Etat « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques

[7] Bien que la circulaire ne s'adresse qu'aux administrations de l'Etat, les collectivités territoriales et les personnes de droit public ou de droit privé chargées d'une mission de service public sont aussi invitées, si elles le souhaitent, à utiliser le portail data.gouv.fr pour mettre leurs données à disposition.

[8] Le site http://publicdata.eu/ recense 205 initiatives dans l'Union Européenne.

Le mouvement n'est que naissant et le terme anglais « open data » encore relativement peu connu du grand public. **L'ouverture des données publiques n'est cependant pas une fin en soi**, mais un moyen de créer de nouveaux services pour les citoyens, d'améliorer le fonctionnement des administrations et de répondre à l'exigence démocratique de transparence de la puissance publique. Ces objectifs ne peuvent être atteints que si les décideurs politiques et les administrations se l'approprient pleinement, et placent la politique des données publiques à un niveau stratégique. La modernisation de l'Etat, en particulier, ne peut être envisagée sans une fluidification et une structuration des échanges entre les administrations et les usagers, et les administrations entre elles. Cela suppose une évolution des cultures, des organisations et des pratiques dépassant largement la seule opération de mise à disposition des données.

Les propositions du CNNum visent à donner une dimension politique et stratégique à l'ouverture des données. Elles poursuivent **trois objectifs** :

- Créer les conditions de la **pérennisation** de l'ouverture des données publiques et, autant que possible, de leur **réutilisation**,
- **Orienter** l'effort sur les jeux de données à **fort potentiel politique, social ou économique**,
- Inscrire l'ouverture des données publiques au cœur de la **modernisation des administrations**.

Elles sont articulées suivant **deux axes d'actions complémentaires** :

- Réformer le droit relatif à la mise à disposition des données publiques : il s'agit d'étendre le champ des informations publiques et de poser des bases solides à leur mise à disposition, en vue d'une réutilisation maximale. Cette réforme pourra s'inscrire dans le cadre de l'élaboration et de la transposition de la nouvelle directive européenne concernant la réutilisation des informations du secteur public.

- Orienter et accompagner la démarche des administrations : les propositions visent à favoriser l'appropriation du mouvement par les administrations et à le développer plus particulièrement autour de projets ambitieux. L'ouverture des données est un levier de modernisation du secteur public car elle favorise la circulation de l'information entre les administrations et la multiplication d'outils pour l'exploiter. Elle amorce une nouvelle conception du service public, répondant directement aux attentes de la société civile et tirant davantage partie des capacités et des initiatives de celle-ci.

Et, afin de permettre leur mise en action :

- Mettre en place une gouvernance des données publiques : l'évolution de la CADA en une Haute Autorité des données publiques doit consacrer le

changement de paradigme amorcé par leur ouverture. Sur le plan opérationnel, il est absolument indispensable d'animer et accompagner les projets, de travailler à s'assurer de la réutilisabilité des données, dans une stratégie globale de modernisation des administrations. Une Agence dédiée pourrait ainsi prendre le relais de la mission Etalab.

Réformer le droit relatif à la mise à disposition des données publiques

Pérenniser le passage à une logique d'offre

Le lancement du portail data.gouv.fr et les initiatives des collectivités territoriales ont impulsé une dynamique importante dont il faut se féliciter. Cependant, le cadre législatif sous-jacent, et notamment la loi du 17 juillet 1978 (dite loi « CADA »), demeure encore par certains aspects, restrictif, ambigu, et inadapté aux caractéristiques et aux enjeux de la société du numérique.

Les modifications apportées au texte (et notamment celles de 2005 et 2011 portant sur la réutilisation des informations publiques) ont été des avancées essentielles, mais elles ont maintenu les nombreuses exceptions, voire en ont introduites de nouvelles pour la réutilisation. Elles n'ont pas élargi le périmètre très restreint de celles qui doivent obligatoirement être publiées ou mises en ligne. Elles ont de manière générale conservé l'esprit de la loi initiale, prévue pour des requêtes en très faible nombre, et rédigée à « l'ère du papier », plusieurs décennies avant l'émergence de la société de la connaissance.

Finalement, la possibilité d'accéder facilement aux données publiques repose encore dans la plupart des cas sur la bonne volonté des administrations, du gouvernement et des collectivités, ce qui n'est pas sans créer de l'incertitude quant à la pérennité de la politique initiée.

Le CNNum recommande une réforme de la loi « CADA » du 17 juillet 1978 visant une diffusion et une réutilisation maximales des données publiques.

La logique qui prévaut dans la loi est encore celle de la demande
La loi « CADA » de 1978 fixe un cadre à l'accès et à la réutilisation des informations publiques. Cependant, la publication (et donc la mise en ligne) n'est requise que pour certains documents, pour l'essentiel législatifs et réglementaires. Les administrations peuvent certes les diffuser (par exemple sur data.gouv.fr) mais dans le cas où elles ne le font pas, l'utilisateur doit s'astreindre à une procédure de demande qui peut prendre jusqu'à 4 mois en cas de refus ou d'absence de réponse (saisine de la CADA).
Cette logique de demande, qui suppose d'ailleurs de l'utilisateur qu'il sache quels documents il souhaite précisément, et la temporalité qui y est associée, ne sont plus adaptées à l'ère du « déluge

des données »[9]. Il faut à cet égard tendre vers une généralisation et une systématisation de la démarche de mise à disposition initiée avec le portail data.gouv.fr et les portails similaires des collectivités. L'offre doit créer la demande et l'innovation plutôt que l'inverse.

Etendre progressivement l'obligation de mise en ligne des informations publiques

Le CNNum propose un élargissement progressif de l'obligation de mise en ligne des informations publiques. Un calendrier pluriannuel, avec certaines priorités (cf. propositions n˚6 et n˚7) permettrait ainsi de laisser le temps aux administrations d'assimiler ce changement de paradigme et de s'y préparer, au niveau technique comme organisationnel.

Sur le plan de la transparence démocratique, une première extension dans les mois qui viennent pourrait par exemple concerner l'intégralité des circulaires, les notes de procédure et notes de services, les marchés et bons de commande. De même, le volet « *data* » bénéficierait d'une mise en ligne obligatoire de bases de données et statistiques (après anonymisation) et de certains jeux de données des EPIC ou services culturels (cf. proposition n˚2 et n˚3). Pour les administrations, cela peut être l'occasion d'actions volontaristes.

Dans le moyen terme, les administrations doivent être assister techniquement et juridiquement dans leur démarche (cf. propositions n˚10 et n˚11).

> **Proposition n° 1 : Accélérer et pérenniser le passage d'une logique de demande à une logique d'offre (mise en ligne) des informations publiques. Etendre progressivement l'obligation de publication des informations publiques, suivant un calendrier pluriannuel.**

Elargir le périmètre des informations publiques

Affiner les exceptions à l'application de la loi du 17 juillet 1978

Par ailleurs, la définition des documents administratifs et des informations publiques exclut encore de nombreuses données. La plupart des exceptions se justifient par l'application d'autres lois relatives à la protection des données à caractère personne, au secret défense, au secret statistique, au secret bancaire, à la propriété intellectuelle etc. Elles n'ont pas ici à être remises en cause, même si certaines pourraient être affinées ou mieux définies, comme le secret des délibérations du gouvernement.

Il pourrait cependant être envisagé de totalement basculer de la notion de «document administratif», qui peut interdire la communication d'un document tout entier dès que certains de ses éléments posent problème, à celle d' «information publique» ou de « donnée publique». C'est déjà le cas des informations relatives à l'environnement[10]. Cet affinage peut requérir un effort plus important des administrations (occultation des éléments concernés), d'où la nécessité de leur fournir

[9] La masse totale de données produites dans le monde augmente d'environ 50% par an ; et autant de données seront produites en 2012 que l'ont été durant toute l'histoire de l'humanité jusqu'en 2009. Plus qu'un changement de volume, c'est un changement de nature.

[10] Article L-124-1 du Code de l'environnement

une assistance juridique et technique (cf. propositions n°10 et n°11), afin notamment d'extraire les informations utiles.

Le droit de réutilisation exclut les données produites par les SPIC

Le potentiel d'innovation de l' « open data » est particulièrement fort en ce qui concerne les données collectées et produites dans l'exercice d'une mission de service public à caractère industriel ou commercial. Les données relatives au transport (horaires, géo-positionnement en temps réel des transports[11] etc.) permettent par exemple à des startups de créer des services à forte valeur ajoutée au bénéfice de tous les usagers, complémentaires au service premier offert par l'entreprise. Ces données pouvant être croisées avec des données d'autres sources, les possibilités d'innovation n'ont comme limite que celle de l'intelligence collective de la communauté des développeurs.

Etendre le champ des informations publiques aux SPIC

Si la loi garantit l'accès à ces données (sauf exceptions prévues par ailleurs), elle les exclut du champ du droit de la réutilisation, empêchant ainsi *a priori* toute possibilité d'innovation par un tiers. Certes, certains EPIC (établissements publics à caractère industriel ou commercial) opèrent, en tout ou partie, dans un milieu concurrentiel. Cependant, la loi prévoit déjà par ailleurs d'exclure du champ des données publiques tout document portant atteinte au secret en matière commerciale ou industrielle. D'ailleurs, l'ouverture des données peut au contraire participer à l'efficacité de ces établissements en créant les conditions de l'innovation ouverte[12]. Le secteur public rattraperait ainsi les entreprises privées, pour la plupart engagées dans ce type de démarche[13].

Le CNNum propose donc une extension du champ des informations publiques et du droit de la réutilisation aux SPIC. Cette mesure résoudrait d'ailleurs certaines incohérences. Dans le cas des transports par exemple, certains EPIC s'opposent à la réutilisation de leurs données alors qu'elles peuvent être par ailleurs des informations publiques, car détenues par les autorités organisatrices du transport public[14]. Le cas des SPIC dont le budget de fonctionnement est lié à la commercialisation des données (ex : Météo-France, IGN) est évoqué à la proposition n°4 portant sur les redevances.

> **<u>Proposition n° 2</u> : Ne prévoir à la définition du périmètre des informations publiques d'autres exceptions que celles découlant de l'application d'autres lois (données à caractère personnel, secret statistique, secret défense etc.). Etendre le droit de la réutilisation aux données brutes des SPIC.**

---

[11] La mise à disposition des données en temps réel requiert une infrastructure particulière. Elle doit donc faire l'objet d'une politique et d'un effort spécifiques de la part de leur détenteur (cf. proposition n°7 sur les projets ouverts).

[12] Cf. rapport « Pour une nouvelle vision de l'innovation », de Paul Morand et Delphine Manceau, rendu au Ministre de l'Economie, de l'Industrie et de l'Emploi en avril 2009 : www.industrie.gouv.fr/enjeux/innovation/rapport-morand-2009.pdf

[13] Certains EPIC se sont d'ailleurs déjà lancés, comme la SNCF : http://data.sncf.com/

[14] L'article L1231-8 du Codes des transports dispose que les autorités organisatrices du transport public de personnes « *instaurent un service d'information, consacré à l'ensemble des modes de transports et à leur combinaison, à l'intention des usagers, en concertation avec l'Etat, les collectivités territoriales ou leurs groupements et les entreprises publiques ou privées de transports* ».

Promouvoir la réutilisation des données culturelles

Des donnés culturelles riches et structurées, propices à la réutilisation

Une grande partie du patrimoine culturel français est aujourd'hui conservée et gérée par des établissements publics, comme les musées, bibliothèques, instituts et centres culturels nationaux, mais aussi les archives. Ces organismes détiennent et produisent une quantité importante de documents et de données, qu'il s'agisse de la numérisation d'œuvres (textes et images), de données de référencement (bases de données, métadonnées, notices descriptives etc.), de données d'activités (horaires et programmations, statistiques de fréquentations, géolocalisation des sites etc.) ou plus généralement de données administratives (budgets, organigrammes etc.). Ces données sont particulièrement riches et structurées grâce aux efforts publics consentis et au savoir-faire développé en matière d'indexation et de catalogage. Elles constituent ainsi un terreau exceptionnellement propice à la réutilisation, comme le montrent des initiatives telles que celle de Bibliothèque nationale de France (data.bnf.fr) ou de la Bibliothèque Nationale et Universitaire de Strasbourg[15].

Une opportunité forte pour le secteur culturel, pourtant encore peu exploitée

Malgré leur potentiel de réutilisation particulièrement fort, les « données culturelles » restent encore peu accessibles, comme en témoigne leur très faible présence sur le portail data.gouv.fr. Permettre le développement d'applications innovantes contribuerait à donner plus de visibilité au patrimoine français, en particulier aux sites et collections les moins connues, et à enrichir l'expérience des visiteurs. Finalement, une politique favorable à la réutilisation des données culturelles serait conforme à la mission du Ministère de la culture de « *rendre accessible au plus grand nombre les œuvres capitales de l'humanité, et d'abord de la France* »[16]. Pour les établissements qui partout dans le monde ont fait ce choix[17], l'ouverture des données est désormais une condition essentielle au rayonnement et à la vitalité de leur patrimoine culturel.

Un régime d'exception difficile à justifier

Le sous-développement en France de l' « open data » des donnés culturelles peut s'expliquer d'une part par le régime d'exception dont elles font l'objet, et d'autre part par des confusions relatives aux questions de propriété intellectuelle.

En effet, par dérogation au chapitre « réutilisation » de la loi du 17 juillet 1978, les établissements, organismes ou services culturels fixent eux-mêmes, le cas échéant, les conditions dans lesquelles les

[15] http://www.bnu.fr/collections/la-bibliotheque-numerique/les-images-de-la-bnu-et-la-licence-ouverte

[16] Décret n°2007-994 du 25 mai 2007 relatif aux attributions du Ministre de la culture et de la communication, porte-parole du Gouvernement. Le rapport « Partager notre patrimoine culturel » (dit « rapport Ory-Lavollée »), remis au Ministre de la culture et de la communication en mai 2010, fait ainsi la recommandation suivante : « *le ministère de la culture doit non seulement accueillir les demandes de réutilisation mais aussi les solliciter par une politique active* ». Mais ce rapport soutient par ailleurs un maintien de « l'exception culturelle. » en matière de réutilisation.

[17] La Harvard Library a ainsi décidé, le 25 avril 2012, de mettre à disposition 12 millions de notices bibliographiques en *Creative Commons Zero* (quasi domaine public). La British Library en avait fait de même fin 2010 (3 millions de notices).

informations figurant dans des documents qu'ils produisent ou reçoivent peuvent être réutilisées (article 11)[18].

Même si un premier élément de jurisprudence a permis d'en lever l'ambigüité[19], cette dérogation demeure un frein majeur à la mise à disposition des données culturelles. L'enjeu n'est pourtant pas celui de la propriété littéraire et artistique puisque l'article 10 de la loi du 17 juillet 1978 précise que « *ne sont pas considérées comme des informations publiques (…), les informations contenues dans des documents sur lesquels des tiers détiennent des droits de propriété intellectuelle* ». La protection de la vie privée a pu être invoquée, notamment dans les cas des fichiers d'Etat Civil détenus par les archives, mais là encore, la loi précise déjà (article 13) que « *la réutilisation d'informations publiques comportant des données à caractère personnel est subordonnée au respect des dispositions de la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés* ». Dans tous les cas, il ne semble pas proportionné de créer un régime d'exception à tout le secteur culturel s'il n'est justifié que par des cas très particuliers, qui restent d'ailleurs à définir précisément. Le CNNum propose donc d'intégrer les données culturelles dans le régime de réutilisation commun. Cette proposition fait écho au projet de directive européenne, qui étend le régime commun de réutilisation aux bibliothèques, musées et archives[20].

Au-delà des données : la confusion demeure sur des questions de propriété intellectuelle

Par ailleurs, des confusions demeurent en pratique sur des sujets de propriété intellectuelle, parmi lesquels la numérisation des documents et le droit d'auteur des agents publics. Ces ambiguïtés créent de l'incertitude aussi bien du côté des établissements culturels que des réutilisateurs au sens large.

Comme le rappelle le Ministère de la Culture et de la communication, les opérations « mécaniques », sans originalité, de numérisation de documents (textes et images) ne « *confèrent à la bibliothèque aucun droit de propriété littéraire et artistique sur les œuvres ainsi reproduites* »[21]. La réutilisation de documents numérisés est pourtant encore fréquemment interdite au nom de la propriété intellectuelle, même lorsque les documents originaux sont tombés dans le domaine public. Mais en présence de droits moraux détenus par des tiers, ces documents ne peuvent pas non plus être qualifiés d'informations publiques, en vertu de l'article 10 cité plus haut.

Une autre « zone grise » concerne le droit d'auteur des agents publics. Ce droit s'applique dès lors que le travail réalisé peut être qualifié d'œuvre de l'esprit. Cela peut être notamment le cas de

[18] Cette dérogation « organique » porte donc également sur des données n'ayant aucun lien avec la création littéraire et artistique, comme les informations administratives, les budgets, les études etc.

[19] Jugement du Tribunal administratif de Clermont-Ferrand rendu le 13 juillet 2011 sur le cas notrefamille.com c/Département du Cantal (dossier n°1001584-1) : « *la dérogation qui leur [aux services culturels] est ouverte par l'article 11 de la même loi ne rétablit pas à leur profit, un pouvoir discrétionnaire d'accorder ou refuser une autorisation de réutilisation* ».

[20] http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/fr.pdf

[21] http://www.culture.gouv.fr/culture/mrt/numerisation/fr/dll/juridi.html#06

certaines notices ou de métadonnées détaillées. Depuis la loi « DADVSI »[22] du 1er août 2006, le droit d'exploitation de l'œuvre est cédé de plein droit, dès la création, à la collectivité publique lorsqu'elle est crée par l'agent « *dans l'exercice de ses fonctions ou d'après les instructions reçues*». Pour autant, cette cession n'est valable que dans la « *mesure strictement nécessaire à l'accomplissement d'une mission de service public* », ce qui exclut *a priori* la réutilisation par un tiers[23]. Par ailleurs, l'existence de droits moraux de l'auteur, même s'ils sont réduits, sort en principe ces données du champ des informations publiques (article 10 de la loi du 17 juillet 1978). En l'absence de décret d'application, les incertitudes demeurent, ou du moins peuvent être invoquées par les établissements producteurs.

Dans ces deux cas, le CNNum recommande l'initiation d'une réflexion sur le cadre juridique en place, et *a minima* une clarification des règles à destination des établissements culturels comme des réutilisateurs.

> **Proposition n° 3 : Intégrer les données culturelles dans le régime de la réutilisation commun. Clarifier les questions de propriété intellectuelle relative à la réutilisation des œuvres numérisées et des données soumises au droit d'auteur des agents publics. Promouvoir l'innovation autour des données culturelles.**

Affirmer le principe de gratuité des données publiques

Les redevances freinent la réutilisation et l'expérimentation

Le droit de réutilisation de certaines données publiques donne aujourd'hui lieu au paiement d'une redevance. Ces redevances mettent en évidence une volonté de valorisation des informations publiques, mais, si elles sont trop élevées, constituent des barrières d'entrée qui sont à l'avantage des gros acteurs privés et favorisent la formation d'oligopoles, au détriment de la diffusion des données et de l'innovation. Elles freinent en particulier l'expérimentation par des startups. Les expériences passées ont montré que la réduction des redevances à des niveaux faibles, voire nuls, pouvaient augmenter le nombre de réutilisateurs de 1000% à 10000%[24]. Les redevances brident dans ce cas le potentiel économique (et indirectement les recettes fiscales qui en résulteraient) et social de la réutilisation des données publiques, pour des revenus souvent marginaux en regard du budget des organismes producteurs.

La gratuité doit s'imposer dans le cas général

Le projet de nouvelle directive européenne[25] prévoit dans le cas général une limite de tarification au recouvrement des « *coûts marginaux de reproduction et de diffusion* » des données. En effet, la collecte et la production des données publiques sont par définition des sous-produits de

[22] Loi n°2006-961 du 1 août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information, modifiant l'article L-131-3-1 du code de la propriété intellectuelle.
[23] Cf. Avis n°20092706 de la CADA
[24] Etude réalisée par Deloitte à la demande de la Commission Européenne, reposant sur une étude de cas issus de toute l'Europe : http://epsiplatform.eu/sites/default/files/models.pdf
[25] http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/fr.pdf

l'accomplissement de la mission de service public, pour laquelle les contribuables ont déjà consenti un effort. De ce point de vue, seuls les coûts spécifiquement liés à leur mise à disposition peuvent justifier une redevance. L'instauration d'une redevance génère cependant des coûts de transaction fixes (personnel, facturation, suivi etc.) qui peuvent s'avérer supérieurs aux revenus ayant vocation à recouvrir les frais de mise en ligne. De plus, les administrations disposent aujourd'hui de l'assistance d'Etalab (ou de la future Agence des données publiques, cf. proposition n°11) et des moyens mutualisés du portail data.gouv.fr pour réaliser ces opérations.

Au moins pour les données brutes, la gratuité s'impose donc en général du point de vue économique et parfois même du point de vue financier pour l'organisme producteur. Le décret du 30 septembre 2005[26] soumet ainsi l'établissement de toute nouvelle redevance à une décision par décret, « *au vu d'éléments dûment motivés*», après consultation du conseil d'orientation de l'édition publique et de l'information administrative (COEPIA).
Cependant, la circulaire du 26 mai 2011 précise que la procédure ne concerne que les redevances instituées postérieurement au 1er juillet 2012. Nous recommandons de la soumettre également aux redevances établies avant cette date, et par ailleurs de limiter la validité de cette redevance à une durée limitée (par exemple 5 ans). A l'issue de ce terme, une nouvelle autorisation devrait ainsi être obtenue.

La priorité demeure cependant la réutilisation des données et le pragmatisme
Cette procédure doit permettre de justifier les redevances et de clarifier leur mode de calcul. Certaines administrations peuvent avoir compté sur ces revenus dans leur budget général, sans corrélation avec le coût de mise à disposition des données. Ces situations n'ont pas vocation à être pérennisées : ces administrations doivent donc être aidées à trouver un financement de remplacement.
D'autres établissements producteurs de données peuvent être soumis à une contrainte d'autofinancement, avoir consenti à des investissements exceptionnels ou être en situation de concurrence. L'existence même des données et leur mise à disposition dans des conditions favorables à leur réutilisation doivent alors primer. Dans tous les cas, le calcul des redevances doit être justifié et viser une réutilisation maximale. Pour les données enrichies, des systèmes non discriminants de partage de la valeur créée peuvent éventuellement être imaginés, en particulier pour les EPIC s'ils venaient à être intégrés dans le régime de réutilisation (cf. proposition n°2).

> **<u>Proposition n° 4</u> : Etendre la procédure de validation des redevances à celles instituées antérieurement au 1er juillet 2012. Ne valider l'établissement de redevances que pour une durée limitée (par exemple 5 ans). Assurer la transparence du calcul des redevances, qui doivent être justifiées.**

[26] Décret n°2005-1755 du 30 septembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n°78-753 du 17 juillet 1978 ; modifié par le décret n°2011-577 du 26 mai 2011.

Clarifier les règles de réutilisation et faciliter le recoupement de données

L'élaboration d'une licence consiste à préciser de manière relativement concise, dans un langage intelligible de tous, les droits et obligations des producteurs comme des réutilisateurs de données. Elle apporte une sécurité juridique aux deux parties, ce qui facilite considérablement la réutilisation des données. Les données du portail de l'Etat data.gouv.fr sont ainsi mises à disposition dans le cadre de la « licence ouverte », qui est également traduite en anglais (« open licence »).

Les cas particuliers, les spécificités de la législation de chaque pays ou des choix politiques divergents peuvent justifier la coexistence de plusieurs licences. Mais une trop grande variété de licences peut être nuisible car les réutilisateurs doivent alors se familiariser avec les subtilités de chacune d'entre elles. Elle défavorise de plus la combinaison de données provenant de sources différentes.

Le mouvement de « open data » est encore relativement naissant. C'est une chance pour les administrations de toute l'Europe qui peuvent dès maintenant entamer des chantiers d'harmonisation. Cela peut aboutir à la proposition d'une famille de licences compatibles entre elles[27] ou à l'émergence d'une licence unique au niveau européen. La Commission européenne a ainsi crée en 2007 une licence de logiciel libre pour les administrations (La Licence Publique de l'Union Européenne), disponible en 22 langues et approuvée par les 27 commissaires. C'est un précédent encourageant.

Le CNNum appelle à la responsabilité des administrations, de l'Etat comme des collectivités territoriales, pour rechercher des consensus et harmoniser autant que possible les licences.

> **Proposition n° 5 : Promouvoir l'émergence d'une licence unique, ou d'une famille de licences compatibles, au niveau européen**

Orienter et accompagner la démarche des administrations

Accentuer l'effort sur des jeux de données à fort potentiel politique, social ou économique

La philosophie première de l'Open Data est de mettre toutes les données publiques à disposition des réutilisateurs, sans faire de distinction a priori. L'idée directrice est de ne pas préjuger du potentiel social et économique d'un jeu de données, car nul ne peut anticiper ce que produira l'imagination collective des développeurs ou ce qu'exigeront de savoir les citoyens. L'offre doit ainsi créer l'innovation et la transparence. Etalab porte donc l'effort sur toutes les données à la fois.

Cette logique, qu'il est nécessaire de conserver, n'est pas suffisante dans la pratique car les administrations n'ont pas forcément intérêt à distribuer les données les plus demandées. De plus, la

[27] La licence ouverte d'Etalab est par exemple compatible avec la « Open Government Licence » (OGL) du Royaume-Uni.

publication de certains documents ou informations peut dans certains cas demander un effort exceptionnellement important (incertitude juridique à lever, anonymisation etc.). La proposition n°1 pose comme point d'horizon la généralisation de l'obligation de mise en ligne des informations publiques. Mais cet objectif, qui requiert une évolution des pratiques et de la culture des administrations, ne sera atteint que progressivement.

Il semble pragmatique, en parallèle, d'accentuer l'effort sur les jeux de données dont la demande, qu'elle vienne de citoyens, de chercheurs ou d'entreprises, démontre *a priori* la valeur sociale ou économique. La version « bêta 2 » du portail data.gouv.fr, lancée le 17 avril 2012, propose ainsi des nouvelles fonctionnalités (« forum » et « boîte à idées ») permettant aux réutilisateurs d'échanger, mais aussi de faire des requêtes précises. La communauté peut ensuite réagir en votant et en commentant ces requêtes. Si les demandes les plus pertinentes ou fréquentes obtiennent effectivement une réponse, c'est un progrès considérable, puisque le demandeur était auparavant contraint à démarcher seul les administrations, voire à saisir la CADA, procédures décourageantes et peu connues du grand public et des startups.

Afin de conforter cette transition, il serait utile de faire connaître auprès des citoyens et des entreprises l'existence de ce dispositif, de s'engager autant que possible sur des délais de réponses et de finalement mettre la priorité sur les jeux de données demandés.

> **Proposition n° 6 : Faire remonter et traiter prioritairement les demandes de données les plus fréquentes ou les plus pertinentes des réutilisateurs. Faire connaître ce dispositif.**

Inscrire l'ouverture des données dans des démarches d'innovation ouverte ciblées

Les administrations doivent dépasser la logique de la simple ouverture des données en lançant des projets ambitieux sur des thèmes précis, qu'ils soient portés par une vision politique, économique ou industrielle. Il peut s'agir par exemple de répondre à un impératif de transparence, d'optimiser des grands systèmes ou de produire des externalités positives pour l'économie.

Cela peut se traduire, suivant les principes de l'innovation ouverte, par des appels à projets, pour lesquels les administrations s'engageraient à fournir toutes les informations nécessaires dans des formes structurées, voire à mettre en place des interfaces de programmation (API). De tels projets favoriseraient le passage d'une logique de stock d'informations à une logique de flux d'informations et la participation de la société civile à la collecte des informations et l'amélioration des services publics (« co-production »). Enfin, il serait très fructueux d'envisager des projets inter-administrations.

Voici quelques idées de projets qui pourraient être ainsi lancés (sans préjuger ici de leur faisabilité juridique et technique), tous reposant sur une exploitation des données des administrations :

- un tableau de bord de l'économie française, en temps réel,

- des algorithmes rapprochant de manière optimale l'offre et la demande pour Pôle Emploi,
- une application permettant de géolocaliser les cabinets des professionnels de la santé et comparer leurs services et honoraires[28].
- une plateforme de signalement et de partage d'informations sur la pollution de l'environnement. Un tel outil permettrait d'optimiser l'action collective (prédictibilité des évènements, mutualisation des dépenses etc.) liée au ramassage des déchets et au traitement des pollutions.
- un site permettant de visualiser le processus d'attribution des logements sociaux et de suivre son propre dossier,
- une carte détaillée des réseaux haut-débit et très haut-débit, détaillant pour chaque zone les technologies et les débits disponibles.

Les administrations peuvent y trouver une occasion de valoriser leur travail, d'accéder à de nouveaux outils facilitant l'accomplissement de leur mission ou, pour certains établissements, d'augmenter leur notoriété et fréquentation.

Proposition n° 7 : Lancer des projets d'innovation ouverte sur des thématiques à forte portée politique, sociale ou économique. Développer à cette fin la mise en place d'API (interfaces de programmation).

Former les agents publics à l'usage et à l'ouverture des données

Le chantier de l'ouverture des données publiques ne peut être mené à bien sans une adhésion et une implication des agents publics. Le CNNum recommande donc de sensibiliser ces derniers aux enjeux de l'ouverture des données publiques mais aussi à l'importance de leur mise en forme, dès la création des données. Cela est d'autant plus nécessaire que les administrations seront les premières bénéficiaires de telles pratiques.

Sensibiliser les agents publics aux enjeux de l' « open data »

Ces actions pédagogiques peuvent prendre la forme d'un volet « open data » dans la formation, initiale ou continue, des fonctionnaires et de séminaires. Dans tous les cas, le thème devrait s'inscrire, de manière centrale, dans le discours sur la modernisation de l'administration. En fluidifiant les échanges d'informations, l'ouverture des données permet en effet de « désiloter » les administrations et de rapprocher les agents publics des usagers, de mettre en évidence les dysfonctionnements et les déséquilibres, mais aussi les bonnes pratiques et les potentiels de mutualisation. Elle place l'administration au cœur de l'innovation et de la vie citoyenne. Elle permet aux agents de récupérer des outils valorisant leur production et donne une nouvelle dimension à leur travail. Elle relativise en effet l'erreur et favorise les raisonnements itératifs.

[28] Ces données peuvent être considérées comme des données à caractère personnel, et ne sont donc pas à proprement parler des données publiques. Elles sont cependant utilisées par le site de l'Assurance Maladie ameli.fr.

Sensibiliser les agents publics sur les aspects juridiques et techniques
Cette pédagogie doit s'accompagner d'une clarification des responsabilités. Les conditions doivent être créées pour que les administrations comme les réutilisateurs aient une idée précise des obligations et des droits de chacune des deux parties, notamment par une clarification du cadre règlementaire (notamment dans le secteur culturel, cf. proposition n°3) et par l'utilisation systématique de licences reconnues. La mise en avant des initiatives les plus avancées peut permettre de dissiper la plupart des craintes.

La sensibilisation doit enfin porter sur l'importance des formats et standards techniques. C'est l'objet de la proposition suivante.

Il ne s'agit pas d'ajouter des tâches aux agents publics mais d'infléchir les pratiques de mise en forme et de description des documents dès leur création. Cependant, ces aspects ne doivent pas constituer des freins à l'ouverture, qui doit rester la priorité. L'assistance d'Etalab (ou de la future Agence des données publiques, cf. proposition n°11) le permet dans tous les cas.

> **Proposition n° 8 : Sensibiliser les agents de la fonction publique aux enjeux et gains de l'ouverture des données publiques et à l'importance des formats. Les former aux démarches à suivre pour assurer leur diffusion.**

Penser la réutilisation des données publiques dès leur création

L'ouverture n'a de sens que si les données peuvent être réutilisées facilement (y compris par l'administration elle-même). Ce principe doit guider toute démarche d' « open data ». Pour être exploitables, les données doivent être structurées, c'est-à-dire lisibles automatiquement par des machines, dans des formats ouverts[29]. Les métadonnées (titre, date de mise à jour, source, mots-clés, géo-référencement) ajoutent enfin beaucoup de valeur aux jeux de données, en facilitant les recherches et classifications.

Il est par ailleurs nécessaire d'effectuer un travail de standardisation des données pour que leurs structures obéissent autant que possible aux mêmes règles quelque soit l'administration qui les produit, et au cours du temps. Cet effort d'harmonisation, qui est la condition d'une interopérabilité large des jeux de données, est particulièrement important dans le cas des collectivités territoriales, si l'on veut pouvoir créer des services à portée nationale ou internationale.

La réflexion entourant la réutilisabilité des données ne se limite d 'ailleurs pas à des questions de formats et de standards. Un exemple important est le niveau de granularité des données diffusées. Des données fines, si possible jusqu'à un niveau individuel (sans que l'identification des personnes, ou des entreprises dans le cas des enquêtes statistiques, ne soit possible) peuvent avoir beaucoup de valeur, notamment pour les journalistes et pour certaines entreprises. Elles permettent par exemple

[29] Les formats exploitables sont recommandés dans le référentiel général d'interopérabilité (RGI)

d'étudier les inégalités et disparités, d'exploiter les distributions géographiques, mais aussi de faire des recoupements avec d'autres données. Cependant, leur diffusion est souvent difficile car la production des données n'a pas été réalisée dans cet objectif. Il faut ainsi penser l'anonymisation dès la collection et la conception des données. Celle-ci peut par exemple être facilitée en évitant de collecter des données personnelles superflues.
Ainsi, l'évolution des pratiques de l'administration pourrait être facilitée par l'élaboration d'un Référentiel Général de Réutilisabilité, sur le modèle des référentiels favorisant l'interopérabilité (RGI) et l'accessibilité (RGAA). Une telle mesure marquerait par ailleurs une nouvelle étape symbolique forte dans l'évolution du service public.

La priorité donnée à la réutilisation des informations publiques pourraient aussi se traduire par l'intégration d'indicateurs de performance appropriés dans le cadre de la LOLF (loi organique relatives aux lois de finances). Les administrations qui assurent le suivi d'établissements publics pourraient prévoir aussi des instruments de suivi et d'évaluation de l'effort des établissements publics en la matière.

> **Proposition n° 9 : Elaborer un Référentiel Général de Réutilisabilité des données publiques, à destination des administrations. Ce référentiel doit notamment prendre en compte les notions de formats, de métadonnées, de standards, et d'anonymisation.**

Mettre en place une gouvernance des données publiques

Les propositions faites précédemment, et les constats qui les sous-tendent, appellent à une évolution de la gouvernance de l'information publique, afin de garantir l'application de la loi et, au niveau opérationnel, d'assister et conseiller les administrations, et faciliter la réutilisation. Plus généralement, il s'agit de fournir aux administrations un cadre et une structure adaptés aux caractéristiques et aux enjeux de la société numérique.

Faire de la politique des données publiques une priorité

La loi du 17 juillet 1978 crée en même temps que le droit d'accès au document administratifs la CADA (commission d'accès aux documents administratifs), autorité administrative indépendante chargée de veiller au respect de ce droit. En 2005, sa compétence est naturellement étendue au droit de réutilisation des informations publiques. Une personne à qui est opposé un refus de communication ou de réutilisation d'informations publiques, et qui souhaite exercer un recours au contentieux, doit obligatoirement saisir au préalable la CADA pour avis.

Ces avis, qui permettent de clarifier le droit, sont cependant consultatifs (l'administration n'est pas obligée de les suivre), et si la CADA est dotée d'un pouvoir de sanction, il ne s'applique qu'aux demandeurs et non aux administrations. Par ailleurs, l'obligation de saisine de la CADA soumet le demandeur à une procédure pouvant durer jusqu'à 4 mois, au terme de laquelle il peut contester la

décision lui faisant grief devant le juge administratif. Cette procédure est d'ailleurs peu connue : la grande majorité des requêtes faites auprès de la CADA porte en fait sur des documents concernant personnellement le demandeur, et les 10% restant rassemblent quelques cas de réutilisation (archives et données relative aux transports).

Le CNNum recommande par ailleurs une évolution de la loi avec le passage progressif à une logique d'offre (mise en ligne) des informations publiques (proposition n°1), et l'élargissement du régime commun aux EPIC et aux établissements culturels (propositions n°2 et n°3), et appelle de manière générale à une évolution des mentalités et des pratiques de l'administration.

A cet égard, le CNNum propose l'évolution de la CADA en **Haute Autorité des données publiques**. Cette autorité serait chargée de :

- veiller au respect du droit relatif aux données publiques, et en particulier d'émettre des **avis conformes**, en cas de saisine suite au refus d'une administration de communiquer des informations, de permettre leur réutilisation, ou, de les mettre en ligne (cf. proposition n°1). Elle serait dotée du **pouvoir d'adresser des injonctions** aux administrations.
- conseiller les administrations qui souhaitent solliciter son avis (comme le fait la CADA).
- exercer une veille afin de promouvoir la transparence et l'ouverture. Elle pourrait ainsi dénoncer les administrations y faisant indûment obstacle (**pouvoir d'autosaisine**) et veiller à assurer le maintien d'un niveau de qualité des données satisfaisant (intégrité des données, mises à jour, niveau de description etc.). Elle aurait à cette fin un **pouvoir d'investigation**.
- proposer éventuellement des modifications des textes législatifs et réglementaires, dans l'objectif d'améliorer la transparence et la diffusion des données publiques (**pouvoir de suggestion**).

<u>Proposition n° 10</u> : Faire évoluer la CADA en Haute Autorité des données publiques, chargée de veiller au respect du droit relatif aux données publiques. Cette autorité administrative indépendante aurait la possibilité d'adresser des injonctions aux administrations, de s'autosaisir et d'enquêter. Elle conseillerait les administrations et pourrait proposer des modifications des textes législatifs et réglementaires.

<u>Accompagner les administrations dans leur transition</u>

Une Agence des données publiques pour accompagner les administrations

La mission Etalab a rendu possible la diffusion d'un grand nombre de données publiques en fournissant aux administrations l'infrastructure, l'impulsion et l'accompagnement nécessaires. Ce soutien opérationnel est absolument indispensable, car les administrations ne disposent pas encore, pour la plupart, des compétences, des incitations et des ressources requises.

Le CNNum recommande donc la pérennisation et l'extension de l'action d'Etalab. La mission pourrait être relayée par une **Agence des données publiques**, en charge de la mise en ligne des données de toutes les administrations, y compris les collectivités territoriales. Ces dernières sont d'ailleurs les

plus susceptibles d'en bénéficier. On compte par exemple plus de 20 000 communes de moins de 500 habitants, qui pourraient produire des informations de proximité potentiellement très intéressantes, notamment pour leurs administrés, mais ne sont pas en mesure de réaliser seules leur « open data ».

Les avantages d'une agence et d'un portail uniques

Un portail unique permet de **mutualiser les coûts de mise à disposition** tout en réduisant les risques de « fracture » dans l'accès aux données. Du côté des demandeurs, la centralisation de toutes les données présente l'avantage de faciliter les recherches, voire de favoriser les « découvertes » inopinées. Cette centralisation n'implique aucune exclusivité puisque les données peuvent être dupliquées et réutilisées. Il est cependant important de veiller à l'interopérabilité entre ce portail et les autres portails des collectivités ou de l'Union Européenne. Il faudrait par ailleurs permettre aux administrations de disposer de leur propre interface, sur leur site, afin qu'elle puisse elles-mêmes communiquer autour de la diffusion de leurs jeux de données et bénéficier des retombées.

Une agence unique offre aussi un interlocuteur visible pour la communauté des réutilisateurs comme pour les producteurs de données, et un espace de dialogue entre les deux parties. Ainsi, Etalab a d'une part mis en place un **réseau de correspondants** dans toutes les administrations, et d'autre part rassemblé l'écosystème de l'innovation autour d'un concours de projets « open data » (« dataconnexions »).

Un accompagnement en amont de la mise à disposition des données

Les propositions faites en deuxième partie de ce document visent toutes à favoriser la mise à disposition de données exploitables, en mettant la priorité sur certains jeux de données, et l'appropriation par les administrations de « l'open data », vecteur de modernisation.

A cet égard, l'Agence des données apporterait comme le fait Etalab une assistance aux administrations, notamment sur la question des formats et l'élaboration des métadonnées. Elle bénéficierait par ailleurs de l'appui et de l'expertise juridique de la Haute Autorité des données publiques, nécessaires par exemple si la mise à disposition de jeux de données demandés par les réutilisateurs (cf. proposition n°6) s'avérait difficile. L'Agence pourrait de plus stimuler et apporter son aide à des projets ambitieux basés sur la réutilisation des données publiques (cf. proposition n°7).

Enfin, elle participerait à la réflexion sur l'évolution des systèmes d'information des administrations, et à l'élaboration du Référentiel Général de Réutilisabilité (cf. proposition n°9) et explorerait les possibilités offerte par la standardisation sémantique des données (web des données ou *linked data*).

Il serait ainsi pertinent d'inscrire la création de cette Agence des données dans un rapprochement entre Etalab, la DISIC et les services « électroniques » de la DGME. Demeure cependant l'inconvénient d'une structure interministérielle : elle ne peut s'adresser aux collectivités territoriales.

<u>Proposition n° 11</u> : Créer une Agence des données publiques, dans le cadre d'un rapprochement d'Etalab, de la DISIC et des services électroniques de la DGME, chargée de faciliter la réutilisation du plus grand nombre de données publiques. Cette Agence doit apporter une expertise technique aux administrations, stimuler et accompagner la mise en œuvre de projets ouverts innovants, et faire remonter les demandes de la société civile. Elle doit enfin contribuer à la mise en place d'une stratégie numérique pour l'administration.

Résumé des propositions

Réformer le droit relatif à la mise à disposition des données publiques

Pérenniser le passage à une logique d'offre

Proposition n° 1 : Accélérer et pérenniser le passage d'une logique de demande à une logique d'offre (mise en ligne) des informations publiques. Etendre progressivement l'obligation de publication des informations publiques, suivant un calendrier pluriannuel.

Elargir le périmètre des informations publiques

Proposition n° 2 : Ne prévoir à la définition du périmètre des informations publiques d'autres exceptions que celles découlant de l'application d'autres lois (données à caractère personnel, propriété intellectuelle et industrielle, secret défense etc.). Etendre le droit de la réutilisation aux données brutes des SPIC.

Promouvoir la réutilisation des données culturelles

Proposition n° 3 : Intégrer les données culturelles dans le régime de la réutilisation commun. Clarifier les questions de propriété intellectuelle relative à la réutilisation des œuvres numérisées et des données soumises au droit d'auteur des agents publics. Promouvoir l'innovation autour des données culturelles.

Affirmer le principe de gratuité des données publiques

Proposition n° 4 : Etendre la procédure de validation des redevances à celles instituées antérieurement au 1er juillet 2012. Ne valider l'établissement de redevances que pour une durée limitée (par exemple 5 ans). Assurer la transparence du calcul des redevances, qui doivent être justifiées.

Clarifier les règles de réutilisation et faciliter le recoupement de données

Proposition n° 5 : Promouvoir l'émergence d'une licence unique, ou d'une famille de licences compatibles, au niveau européen

Accompagner et orienter la démarche des administrations

Accentuer l'effort sur des jeux de données à fort potentiel politique, social ou économique

Proposition n° 6 : Faire remonter et traiter prioritairement les demandes d'informations les plus fréquentes ou les plus pertinentes des réutilisateurs. Faire connaître ce dispositif.

Inscrire l'ouverture des données dans des démarches d'innovation ouverte ciblées

Proposition n° 7 : Lancer des projets d'innovation ouverte sur des thématiques à forte portée politique, sociale ou économique. Développer à cette fin la mise en place d'API (interfaces de programmation).

Former les agents publics à l'usage et à l'ouverture des données

Proposition n° 8 : Sensibiliser les agents de la fonction publique aux enjeux et gains de l'ouverture des données publiques et à l'importance des formats. Les former aux démarches à suivre pour assurer leur diffusion.

Penser la réutilisation des données publiques dès leur création

Proposition n° 9 : Elaborer un Référentiel Général de Réutilisabilité des données publiques, à destination des administrations. Ce référentiel doit notamment prendre en compte les notions de formats, de métadonnées, de standards, et d'anonymisation.

Mettre en place une gouvernance des données publiques

Faire de la politique des données publiques une priorité

Proposition n° 10 : Faire évoluer la CADA en Haute Autorité des données publiques, chargée de veiller au respect du droit relatif aux données publiques. Cette autorité administrative indépendante aurait la possibilité d'adresser des injonctions aux administrations, de s'autosaisir et d'enquêter. Elle conseillerait les administrations et pourrait proposer des modifications des textes législatifs et réglementaires.

Accompagner les administrations dans leur transition

Proposition n° 11 : Créer une Agence des données publiques, dans le cadre d'un rapprochement d'Etalab, de la DISIC et des services électroniques de la DGME, chargée de faciliter la diffusion et la réutilisation du plus grand nombre de données publiques. Cette Agence doit apporter une expertise technique aux administrations, stimuler et accompagner la mise en œuvre de projets ouverts innovants, et faire remonter les demandes de la société civile. Elle doit enfin contribuer à la mise en place d'une stratégie numérique pour l'administration.

Remerciements

La rédaction de cet avis s'est appuyée sur un groupe de travail extérieur au CNNum, dirigé par **Jérôme Cukier**, auquel ont participé notamment François Bancilhon, Nadia Barigault, Mehdi Benchoufi, Denis Berthault, Pierre Chzranowski, Emmanuel Condamine, Claire Gallon, Jean-Baptiste Ingold, Benjamin Jean, Sébastien Le Bail, Emmanuel Marcovitch, Maurice Ronai.

Rapporteur général (CNNum) : Laurent Guérin

Annexe I – Tour d'horizon de projets « Open Data » à travers le monde

Les quelques exemples qui suivent dépassent le seul périmètre des données publiques mais ont vocation à illustrer concrètement les possibilités offertes par l'approche « *open data* », en termes d'innovations, d'améliorations de la qualité du service public et d'avancées démocratiques.

Transport

Le transport connaît un foisonnement de projets, permettant d'améliorer les déplacements et de transformer l'expérience de la ville. Citons entre autres : Open Plan Project, permettant de connaître le trafic des transports publics en temps réel, Open trip planner, ou Exit Strategy qui promettent de faire gagner plusieurs minutes sur n'importe quel trajet en transport en commun à New York. Roadify, une application permettant d'échanger des informations en temps réel avec la communauté des usagers des transports, a également connu un succès fulgurant. Be-part propose quant à elle de surveiller la bonne conduite des projets publics de transport. San Francisco propose une grande variété de données de transport (sécurité, wifi, aménagement des voiries, météo etc.) tout comme la ville de Vancouver au Canada. En France, l'opérateur de transport public de Rennes met ses données à la disposition des développeurs.

Collaborations scientifiques

Le taux de diffusion et de citation augmente largement lorsque les informations scientifiques sont en accès libre. L'enjeu de l'ouverture des données est tel qu'un certain nombre d'éditeurs, comme la revue Nature, imposent le partage des données comme condition de publication. Certaines agences publiques de recherche mettent à la disposition toutes leurs données afin de nourrir et stimuler les travaux d'équipes tierces.

Politique et démocratie

En Angleterre, le projet Public Whip suit l'activité parlementaire des élus, et publie notamment les donnés publiques relatives à leurs votes et débats. De manière similaire, l'organisation Maplight examine de près les campagnes des candidats aux différentes élections, à la recherche d'irrégularités et de conflits d'intérêts.

Aux Etats-Unis, Open Corporate croise les contrats publics et l'activité de lobbying de certaines entreprises pour mettre en évidence les éventuels conflits d'intérêt.

Le site de l'Utah est un exemple emblématique de transparence des dépenses publiques. Le citoyen peut se faire une idée particulièrement précise de l'emploi des fonds publics, secteur par secteur. OpenSpend permet la même chose au niveau national. Dans le même esprit, a été développé le NYC City Map, qui a permis la visualisation des investissements territoriaux de la ville. La ville de Chicago a publié l'ensemble des contrats publics relatifs à la passation de marchés publics et ce depuis 1993.

La Suède a mis en ligne le site OpenAid pour documenter tous ses efforts d'aide publique au développement.

Parmi les pays s'étant dotés d'un portail open data, certains sont emblématiques de la nature mondiale du mouvement, comme le Kenya ou les Émirats Arabes Unis.

Santé

En Australie, de très nombreuses informations relatives au système de santé sont mises en ligne, comme par exemple My Hospitals qui permet de comparer les performances des hôpitaux. Au Royaume-Uni, les indicateurs de santé mis en place par le National Health Service permettent aux usagers de cibler leur choix de soin, ou encore de construire leurs propres indicateurs de qualité du système de soins. La plupart des hôpitaux britanniques affichent leurs indicateurs de performances (KPI)[30].

Réseaux intelligents : partage des données et puissance de calcul

La Commission Européenne a porté un projet ambitieux, European Grid, qui consiste à mutualiser les données relatives aux infrastructures et la puissance de calcul des ordinateurs associés. Ont été ainsi rendues possibles des avancées remarquables en climatologie ou sur le développement de nouvelles énergies. Le partage des données est ainsi une ambition affichée de la Commission Européenne : l'ambition étant d'organiser le territoire « infostructurel » qui permettra le partage des données dans de nombreux domaines : Enabling grids for e-Science (science dures), NMDB (astrophysique), Impact Project (biochimie), Parse Insight (préservation des données). De la même façon, aux Etats-Unis, en Grande-Bretagne (UKRDS) ou en Australie (ANDS), les grandes agences de recherche publique ont pour objectif d'établir une « infostructure » dédiée au partage des données, de leur traitement, de leur organisation et de leur visualisation.

Sciences humaines

Le projet européen « Dariah Project » ouvre de nouvelles perspectives en histoire, archéologie etc. Il permet par exemple l'analyse dans le temps et dans l'espace de l'évolution de phénomènes archéologiques.

Opencontext permet de mettre à jour et de partager les découvertes archéologiques.

En sciences sociales, de nombreux pays, comme la Norvège, investissent le champ de l' « open data » afin de mieux comprendre leur démographie, la structure de la consommation selon les revenus ou encore les phénomènes d'inégalités.

Concours

De nombreux concours sont organisés par les institutions ouvrant leurs données : « Apps for Democracy » à Washington D.C., « Big Apps Challenge » à New York, « Apps for Metro » à Chicago. Au Royaume-Uni, L' « Open Knowledge Foundation » et l' « Open Forum Academy » organisent aussi des concours d'applications innovantes, tel que l'Open data Challenge.

[30] http://www.barnsleyhospital.nhs.uk/annualreport/director-report/key-performance-indicators-kpis/