

Rapport sur l'open data

Samuel Goëta (Dataactivist) pour le Conseil départemental de l'Ardèche

2018-01-22

Contents

Introduction

Ce rapport vise à expliquer les essentiels d'une démarche d'ouverture des données et à décrire les facteurs de succès d'un projet d'open data. C'est un document à visée pédagogique qui, sans entrer dans les détails du contexte institutionnel et des spécificités de chaque jeu de données, permet d'établir un niveau commun de compréhension pour les acteurs de l'ouverture des données en Ardèche. Il repose sur des sources publiques qui sont référencées dans le document, elles sont autant de prolongements pour cette introduction à l'ouverture des données.

Dans la première partie, le rapport présente les grands principes internationaux de l'ouverture des données et retrace quelques uns des jalons de l'émergence de l'open data. On y découvre quelques uns des acteurs à l'origine de l'ouverture des données et les principes qui y ont été formulés.

La deuxième partie aborde le contexte juridique de l'ouverture des données en France. Elle remonte aux origines de l'open data dans le droit d'accès et de réutilisation des informations publiques pour montrer la continuité avec le principe d'ouverture des données par défaut prévu par la loi pour une république numérique.

La troisième partie s'intéresse aux bénéfices organisationnels, informationnels, politiques et économiques de l'ouverture des données. Elle insiste sur les bénéfices immédiats d'une démarche d'open data sans se concentrer uniquement sur le volet économique comme on peut le lire souvent dans la littérature.

Enfin, la quatrième partie présente quelques uns des facteurs de succès d'une politique d'open data : obtenir un soutien au plus haut niveau, configurer l'organisation pour l'ouverture, faciliter la découverte et l'utilisation des données, rester à l'écoute et interagir avec les usagers.

Ce rapport est une première version, n'hésitez à nous faire parvenir vos retours à l'adresse suivante : samuel@dataactivi.st.

Chapter 1

Les grands principes d'une démarche d'open data

Comme nous allons le voir dans cette première partie, le cadre juridique de l'ouverture des données repose sur des racines anciennes mais l'open data en tant que tel est apparu récemment, il y a moins de 10 ans, avec des grands principes qui se sont consolidés avec le temps. Dans cette première partie, nous allons revoir ensemble les grands principes de l'open data : leurs origines, leur adaptation en France et les bénéfices pour une collectivité à les adopter.

1.1 Aux origines de l'ouverture des données : retours sur quelques grandes dates fondatrices

En cinq épisodes, revenons sur les principaux moments de définition des grands principes internationaux de l'open data. Cette partie permettra de replacer l'open data dans son contexte d'apparition, de mieux connaître les acteurs à l'origine des grands principes et de comprendre les textes de référence de l'ouverture des données. Nous résumerons les grands principes de l'ouverture des données dans la partie suivante.

Avant de revenir sur ces différents épisodes, il faut rappeler que le terme d'open data a des origines plus anciennes que la dernière décennie sur laquelle nous allons nous concentrer ici. Le terme est apparu pour la première fois dans les années 1970 dans les accords qu'a signés la NASA avec des pays partenaires en vue du partage de données satellitaires. C'est en 1995 qu'on en voit le premier usage public aux Etats-Unis dans un rapport de la National Academy of Science intitulé *On the Full and Open Exchange of Scientific Data*.

1.1.1 2005 : Open Definition, la définition juridique des droits de l'utilisateur du savoir ouvert

En août 2005, le chercheur en économie Rufus Pollock, fondateur de l'Open Knowledge Foundation (OKFN), une organisation à but non lucratif qui vise à "promouvoir l'ouverture de toutes les formes de savoir", invitait les premiers membres de l'OKFN et son réseau de partenaires à adopter collectivement une définition du savoir ouvert. Dans son appel à commentaire (*Request for Comments*), Pollock souhaitait décliner une série de conditions essentiellement juridiques permettant d'établir qu'un savoir est ouvert. La définition devait aussi servir à énumérer les licences ouvertes spécifiques au savoir et à fédérer des disciplines éparses.

Cette définition se fonde directement de l'expérience du mouvement de l'*open source*, l'ouverture du code informatique, une généalogie clairement affirmée dans le texte de l'Open Definition qui crédite l'*Open Source Definition* comme la ressource essentielle qui a servi à la rédaction de la définition mais aussi à forger



Figure 1.1: La photo de famille des participants de la rencontre de Sebastopol

l'idée même d'ouverture. Cet effort de définition s'inscrit aussi dans le prolongement du travail de Creative Commons qui a défini une série de licences assorties à des droits et devoirs des usagers d'un savoir ouvert.

Pour la résumer en quelques mots, l'Open Definition (quelques années après sa publication) décline les conditions de l'ouverture du savoir. Cette définition utilise la notion de savoir pour désigner un domaine très large, qui rassemble des objets informationnels très différents (donnée, document, contenu, œuvre, article...) Sans entrer dans le détail de chacune des clauses, l'Open Definition exclut les licences qui « discriminent » selon les types d'usagers ou la finalité de la réutilisation. Elle demande d'accorder trois droits fondamentaux (utiliser, réutiliser, redistribuer) et autorise à contraindre les réutilisateurs à deux exigences possibles : la citation de la source et le partage des modifications de l'œuvre avec la même licence (clause de share alike).

En posant la base d'un élargissement de l'*open source* au savoir, l'Open Definition a constitué une ressource précieuse pour l'ouverture des données publiques. Elle a établi des critères essentiellement juridiques qui caractérisent l'ouverture en termes de droits des usagers sans préjuger du type de savoir concerné. Cet effort de définition s'est inscrit aussi dans le prolongement du travail de Creative Commons qui a défini une série de licences assorties à des droits et devoirs des usagers d'un savoir ouvert.

Aller plus loin : lire l'Open Definition.

1.1.2 2007 : la rencontre dite de Sebastopol, la définition des grands principes de l'open data

Après avoir découvert les fondements juridiques de l'ouverture des données avec l'Open Definition, nous partons maintenant aux Etats Unis à la rencontre de ceux qui ont défini les principes encore aujourd'hui en vigueur de l'open data.

Le 22 octobre 2007, une invitation est envoyée aux membres d'un groupe de travail sur l'Open Government pour une rencontre les 7 et 8 décembre 2007 à Sebastopol en Californie au sein des locaux de la maison d'édition O'Reilly. Les organisateurs de cette rencontre sont Carl Malamud qui dirige le site associatif PublicResource.org et Tim O'Reilly, le directeur de la maison d'édition O'Reilly spécialisée dans les sujets technologiques et l'édition électronique ouverte. Dans le texte de l'invitation, les deux organisateurs se sont fixés pour ambition de lister dix principes de l'*open government* afin que les candidats à l'élection du président des États-Unis suivent leurs recommandations. Trente participants sélectionnés par les organisateurs en fonction de leur affiliation à une organisation qui demande, ouvre ou réutilise des données ont accepté l'invitation de Malamud et O'Reilly.

Au termes des deux jours de la rencontre de Sebastopol, les participants ont défini ensemble une série de huit critères pour que des données gouvernementales soient considérées comme ouvertes : - **complètes** : toutes les données publiques doivent être rendues disponibles dans les limites légales liées à la vie privée ou la sécurité ;

1.1. AUX ORIGINES DE L'OUVERTURE DES DONNÉES : RETOURS SUR QUELQUES GRANDES DATES FONDATR

- **primaires** : les données ouvertes sont telles que collectées à la source, non-agrégées avec le plus haut niveau de granularité ;
- **fraîches** : les données doivent être disponibles dès qu'elles sont produites (*timely*) ;
- **accessibles** : les données doivent être utilisables par le plus grand nombre d'utilisateurs potentiels ;
- **lisibles par les machines** : les données peuvent faire l'objet d'un traitement automatisé par les machines ;
- **non discriminatoires** : elles peuvent être utilisées par tous sans réclamer un enregistrement préalable ;
- **dans un format ouvert** : ce format ne doit pas être la propriété d'une organisation en particulier et doit faire l'objet d'une gouvernance commune par ses usagers ;
- **avec une licence ouverte** : les principes de Sebastopol vont plus loin que l'Open Definition en demandant que les données soient placées dans le domaine public.

Ces principes sont aujourd'hui encore le fondement de l'open data. Les participants de la rencontre de Sebastopol ont rempli leur objectif, à savoir l'adoption de ces principes par le futur président des États-Unis puisque le 21 janvier 2009, jour de son investiture à la Maison-Blanche, Barack Obama a signé deux mémorandums sur l'*Open Government*. Le premier exigeait une plus grande coopération des agences gouvernementales aux procédures du *Freedom of Information Act* (FOIA). Le second réclamait que les agences gouvernementales mettent en œuvre des politiques en faveur de la transparence, la collaboration avec la société civile et la participation des citoyens qui a abouti au lancement en 2009 de data.gov, le premier portail open data national.

Aller plus loin : le site OpenGovData.org propose une version annotée des principes de Sebastopol (en anglais).

1.1.3 2008 : “Raw Data Now”, l'appel du fondateur du web à l'ouverture des données brutes

Tim Berners-Lee, l'inventeur du web, a formulé son appel à l'ouverture des données brutes le 4 février 2009 à Long Beach en Californie lors d'une conférence TED. TED est un réseau de conférences retransmises gratuitement sur le web qui vise à présenter simplement des idées et à convaincre l'audience de s'impliquer.

Dans la vidéo de la conférence dépassant aujourd'hui le million de vues, Tim Berners raconte d'abord son parcours au sein du CERN, l'accélérateur de particules, où il a développé le web pour faciliter le partage des documents produits dans son laboratoire. Berners-Lee dit ressentir la même difficulté pour accéder aux données qu'à l'époque de la création avec les documents. Pourtant, les données déterminent une grande partie de nos vies. Il se félicite de la naissance de l'open data et des engagements pris par le président Obama à son arrivée à la Maison Blanche (son discours est intervenu deux mois après la signature des mémorandums) mais il estime que l'ouverture des données implique aussi de transformer les attitudes des administrations.

Pour l'inventeur du web, très souvent, les agents publics sont tentés de garder leurs données et trouvent une multitude de raisons pour ne pas les diffuser et permettre leur réutilisation. Dans sa présentation, Berners-Lee a fait référence au médecin suédois Hans Rosling qui a proposé l'expression « *database hugging* », une métaphore pour décrire une attitude dans laquelle les agents de l'administration s'accrochent à leurs données au point de les « câliner ». Berners-Lee a repris cette métaphore et l'a mimée sur la scène de TED.

Pour l'inventeur du web, les administrations n'arrêtent le *database hugging* qu'à partir du moment où elles ont présenté leurs données sur un beau site web. Il a demandé d'inverser cette logique et d'abord de fournir les données.

Hans appelle ça le *database hugging*. Vous serrez votre base de données. Vous ne la laissez pas partir tant que vous n'en avez pas fait un joli site web. [...] Faites-en donc un joli site. Mais avant



Figure 1.2: Tim Berners-Lee, lors de sa conférence TED de 2009, mimant le **database hugging**, l'attitude des administrations qui « s'accrochent » à leurs données

cela, donnez-nous accès aux données non altérées. On veut des données. On veut des données non altérées. **Il faut que nous demandions des données brutes maintenant.**

Tim Berners-Lee demande alors au public de la conférence TED de crier « *Raw data now!* » (« Des données brutes maintenant ! ») à l'attention des administrations (figure 1.3).

Ce discours de Tim Berners-Lee a imposé la demande de données brutes comme un aspect essentiel de l'*open data* avec un slogan facilement mémorisable : ouvrez les données brutes maintenant ! Cette demande de données brutes s'explique par deux choses. D'une part, en ouvrant les données telles qu'elles sont produites, les administrations n'auraient pas à les retravailler, ce qui a été pensé comme un levier pour faciliter l'ouverture. D'autre part, l'obtention des données brutes est pensée comme un moyen de réduire les asymétries d'information entre l'administration et la société civile puisque les données brutes seraient le matériau de l'information publique avec son traitement par l'administration.

Aller plus loin : la vidéo de la conférence de Tim Berners-Lee est traduite en français sur le site de TED.

1.1.4 2010 : le modèle en 5 étoiles, une échelle de l'ouverture des données

Après avoir exigé l'ouverture des données brutes, Tim Berners-Lee a appelé à l'utilisation de formats ouverts de données. En 2010, il propose un modèle en cinq étapes, une hiérarchie de la première à la cinquième étoile qui, à la manière de la classification des hôtels, permet aux réutilisateurs de distinguer la qualité des données. Ce modèle s'adressait particulièrement aux gouvernements pour les encourager à adopter le Linked Data pour ouvrir leurs données.

Sur la boutique en ligne du W3C, le consortium en charge des standards du web, Berners-Lee vend même des tasses sur lesquelles figure son modèle en cinq étoiles. Il a déclaré espérer que la circulation de ces tasses dans les bureaux inciterait à ouvrir et lier toujours plus de données.

Dans la hiérarchie de Tim Berners-Lee, les données sont ouvertes dès la validation du premier critère. Il considère que, plus une donnée obtient d'étoiles, plus elle sera simple à utiliser :

- la première étoile demande la publication sur le web des données, quel que soit leur format avec une licence ouverte.



Figure 1.3: Tim Berners-Lee, lors de sa conférence TED de 2009, Tim Berners-Lee appelle le public à crier « *raw data now* »



Figure 1.4: Tasse du W3C reprenant le modèle en cinq étoiles de Tim Berners-Lee.



Figure 1.5: Session de travail des membres du G8 en 2013

- la deuxième étoile exigeant que les données publiées sur le web sous une licence ouverte soient lisibles par les machines et structurées.
- la troisième étoile réclame la publication des données dans un format non propriétaire.
- pour obtenir la quatrième étoile, les données doivent être publiées dans les standards ouverts du W3C (RDF et SPARQL) qui imposent que les objets contenus dans les données soient décrits.
- la cinquième étoile demande qu'elles soient liées à d'autres données publiées sur le web.

Dans les projets d'open data, le modèle de Tim Berners-Lee a été employé essentiellement pour inciter les agents à ouvrir les données dans des formats ouverts comme le CSV plutôt que d'utiliser le format Excel. L'utilisation de formats sémantiques, les deux derniers niveaux du modèle, réclame un travail trop important de transformation des données au regard des moyens généralement alloués aux projets d'*open data*.

Retenons donc du classement en cinq étoiles qu'il suggère aux administrations d'ouvrir les données de manière progressive. En quelque sorte, il leur propose une marche à suivre : d'abord publier les données sur le web avec une licence ouverte, ensuite avec des formats lisibles par les machines puis dans des formats ouverts et enfin éventuellement selon les standards du Linked Data.

Aller plus loin : le site 5stardata.info présente en détail le modèle cinq étoiles de Tim Berners-Lee.

1.1.5 2013 : la charte internationale de l'open data, vers l'ouverture par défaut

Les 17 et 18 juin 2013 à Loughe-Erne en Irlande du Nord, le Premier ministre britannique, David Cameron, accueillait la réunion du G8, la rencontre de huit chefs d'État parmi les plus grandes puissances économiques mondiales (Allemagne, Canada, États-Unis d'Amérique, France, Royaume-Uni, Italie, Japon, Russie). L'agenda comportait une session sur l'adoption d'une charte sur l'open data.

La charte sur l'open data du G8 a été publiée à la suite du G8 et se compose d'une série de cinq principes et trois annexes. Son préambule détaille les bénéfices de l'open data : création de services, transparence de l'action publique, meilleure gouvernance, amélioration du débat public, lutte contre la corruption, soutien à l'innovation des entreprises et de la société civile, prospérité renouvelée...

Cette charte a été par la suite reprise par un groupe qui s'est notamment réuni lors de l'Open Data Conference de 2015 à Ottawa afin de produire une charte internationale de l'open data, dépassant les seuls pays du G8, qui synthétise les grands principes de l'ouverture des données. Elle a été officiellement publiée en octobre 2015 en marge de l'Assemblée Générale des Nations Unies.

La charte internationale de l'open data fixe cinq grands principes pour l'ouverture des données :

- le premier principe donne pour objectif de **faire de l'open data la pratique par défaut des administrations** pour les données publiques tout en respectant les législations en vigueur sur la propriété intellectuelle et la vie privée.
- le deuxième principe demande la publication des données **en temps opportun (*timely*) et de manière exhaustive**, c'est-à-dire en fournissant des données désagrégées et de qualité.
- le troisième principe réclame **des données accessibles et utilisables** fournies dans des portails centraux, sans enregistrement préalable, avec une licence ouverte.
- le quatrième principe demande des **données comparables et interopérables** fournies dans des formats structurés et normalisés favorisant l'interopérabilité et la réutilisation.
- le cinquième principe fixe pour objectif à l'open data d'**améliorer la gouvernance et d'encourager la participation citoyenne** dans la logique de l'ouverture des gouvernements.
- le sixième principe vise à **favoriser le développement inclusif et l'innovation** du fait que les données ouvertes donne du pouvoir d'agir à tous.

À travers ce résumé, on voit donc que la charte internationale de l'open data s'inscrit dans la continuité des définitions de l'open data évoquées précédemment. Elle fixe des grands objectifs politiques qui dépassent les textes évoqués précédemment. De ce fait, elle constitue une ressource moins opérationnelle pour définir si une donnée est ouverte ou non.

Aller plus loin : lire la charte internationale de l'open data en français sur le site [OpenDataCharter](#)

1.2 Revue des grands principes de l'ouverture des données

Dans la section précédente, nous avons vu les grands moments lors desquels l'open data a été défini. Nous allons maintenant résumer les principes de l'open data en reprenant les dix points définis par la Sunlight Foundation. Il n'y a pas de différence majeure avec les définitions de l'open data évoquées précédemment mais la Sunlight Foundation a réussi à synthétiser l'essentiel dans ce document.

Les données doivent être :

1. **complètes** : elles doivent représenter l'intégralité de ce qui est collecté par l'administration sur un sujet donné avec des métadonnées qui explique comment elles ont été collectées, c'est le **principe d'open data par défaut** réclamé par la charte internationale de l'open data qui est aujourd'hui la norme en France avec la loi pour une République Numérique ;
2. **primaires** : les données doivent être non agrégées et telles que collectées à la source, cela équivaut à la demande de **données brutes** popularisée par Tim Berners-Lee ;
3. **fraîches** : les données doivent être publiées dès que possible pour garder leur valeur ;
4. **accessibles** : les données doivent être directement téléchargeables sans enregistrement préalable ;
5. **lisibles par les machines** : les données doivent être exploitables automatiquement ;
6. **accessibles sans discrimination** : les usagers doivent pouvoir accéder aux données à tout moment sans s'enregistrer ni fournir de justification ;
7. **disponibles sous des formats ouverts** : les spécifications de standards de données doivent être ouvertes et, si possible, faire l'objet d'une gouvernance partagée ;
8. **disponibles sous licence ouverte** : l'Open Definition donne les critères juridiques d'une licence ouverte ;
9. **accessibles de façon pérenne en ligne** : les données doivent être archivées et les versions précédentes conservées ;
10. **sans coût d'utilisation** : l'accès et la réutilisation des données doivent être gratuits.

Aller plus loin : lire les dix principes définis par la Sunlight Foundation.

Chapter 2

Le cadre juridique de l'open data en France

On pourrait croire que l'open data constitue un mouvement récent né il y a près de dix ans. En réalité, les principes internationaux que nous venons de découvrir s'inscrivent dans la continuité du droit français et européen qui garantit la transparence de l'action de l'administration. En plus de porter l'attention sur les données brutes de l'administration, les principes de l'open data ont apporté deux nouveautés : la consécration du droit de réutilisation avec l'apparition de standards internationaux et l'importance de la lisibilité des données par les machines.

2.1 La loi CADA : accorder aux citoyens un “droit de savoir”

On peut retracer la philosophie de l'open data dans les fondements même de la République avec la **Déclaration des Droits de l'Homme et du Citoyen de 1789** qui dispose dans son article 15 que “la Société a le droit de demander compte à tout Agent public de son administration.” Cet article pose un devoir de reddition de comptes pour l'administration qui s'est exercé pendant longtemps principalement par des institutions comme la Cour des Comptes en charge du contrôle de l'action public.

Néanmoins, c'est véritablement à partir de 1978 qu'a émergé un droit d'accès à l'information publique qui s'est progressivement transformé en droit de réutilisation avec le développement d'Internet. Le droit d'accès des citoyens à l'information publique émerge en **1978 avec la loi dite CADA** du nom de la Commission d'Accès aux Documents Administrations Administratifs. La France était alors le troisième pays au monde après la Suède en 1766 et les Etats-Unis en 1966 avec le Freedom of Information Act (FOIA) en 1966 à accorder un “droit de savoir” à ses citoyens qui permet de demander aux administrations les informations publiques qu'elles détiennent. En 1978, c'est **un droit d'accès** qui est accordé aux citoyens, la priorité à l'époque étant d'améliorer la relation entre l'administration et le public et de renforcer la transparence de l'action publique.

La loi CADA donne **une définition très large d'un document administratif** comme “tous les documents produits ou reçus par l'administration qu'ils se présentent sous forme écrite (dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires...), sous forme d'enregistrement sonore ou visuel ou sous forme numérique ou informatique.” C'est dans ce cadre très générique que s'inscrit l'ouverture des données. Le droit d'accès s'applique aux documents administratifs **produits dans le cadre d'une [mission de service public](de manière directe par les acteurs publics ou indirecte (délégation de service public))**, c'est-à-dire l'accomplissement de toute activité prise en charge de manière directe par les acteurs publics ou indirecte (délégation de service public) pour satisfaire l'intérêt général et régi par des principes de continuité, d'égalité et d'adaptabilité. Enfin, le droit d'accès ne s'exerce que si l'administration a effectivement en sa possession le document demandé, que si le

document est formellement **achevé** et qu'il n'est pas **préparatoire à une décision administrative** en cours d'élaboration.

Sont **exclus** du droit d'accès les documents administratifs pouvant atteindre porter atteinte à l'exercice des activités régaliennes de l'État et à l'intérêt général. Plusieurs secrets sont protégés notamment concernant les délibérations du Gouvernement, le secret de la défense nationale, la sécurité publique ou encore le déroulement des procédures juridictionnelles. Surtout, le cadre de la loi CADA exclut les données à caractère personnel, relatives à des personnes physiques et permettant leur identification directe ou indirecte. Toutefois, les documents comprenant des données à caractère personnel ne sont communicables et réutilisables que si au moins l'une des trois conditions suivante est remplie : 1. le consentement des personnes concernées a été recueilli après leur bonne information sur la finalité et les modalités de la communication ou de la réutilisation des données les concernant ; 2. les données sont anonymisées (c'est-à-dire non identifiantes ou ne permettant pas, compte tenu de leur niveau d'agrégation, d'identifier à nouveau les personnes concernées) ; 3. la réutilisation est autorisée par un texte législatif ou réglementaire.

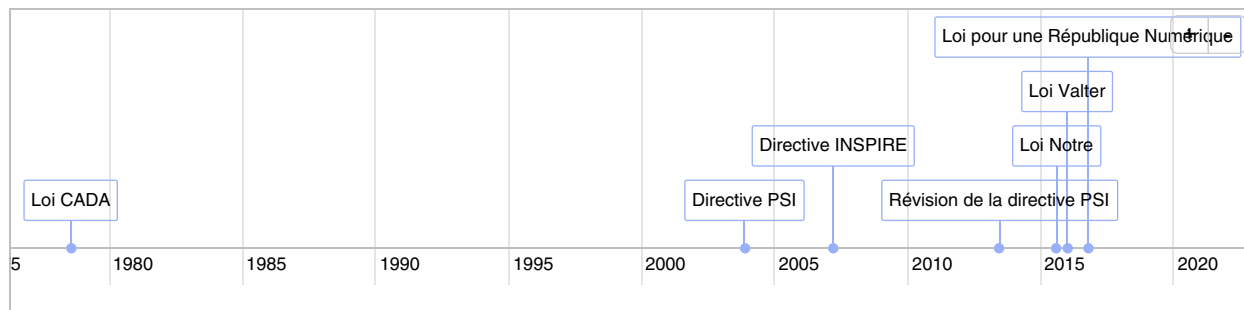
Aller plus loin : Sur tous ces points, le site de la CADA détaille la jurisprudence, en particulier dans la partie sur l'étendue du droit d'accès. Le guide de demande de données publiques auprès des collectivités produit par Bordeaux Métropole permet aussi de mieux comprendre le cadre légal.

2.2 Du droit d'accès au droit de réutilisation

La loi CADA a accordé aux citoyens un droit d'accès aux documents administratifs, la question de la réutilisation des données publiques a émergé dans les années 1990 avec le développement d'Internet facilitant la diffusion des données publiques et le développement de nouveaux services au public. Comme l'explique Boustany (2013), « la différence réside dans les enjeux : l'accès aux documents publics relève de la démocratie et de la transparence des institutions ; la réutilisation des données des données pour produire des services ou pour la connaissance présente des enjeux socio-économiques. »

Aller plus loin : lire l'article de Joumana Boustany (2013), « Accès et réutilisation des données publiques : État des lieux en France » dans la revue *Les Cahiers du numérique*

La frise qui suit retrace les principales dates de l'apparition du droit de réutilisation des données, chacune de ces dates étant développée à la suite :



Le **droit de réutilisation des informations publiques** a été fortement soutenu par l'Union Européenne en particulier avec la **Directive dite PSI (*Public Sector Information*) de 2003** concernant la réutilisation des informations du secteur public qui fixe les règles minimales et les conditions de réutilisation de ces données publiques. Elle a été transposée en droit français par l'ordonnance du 6 juin 2005 qui autorise les usages commerciaux des données publiques et fixe le cadre de la réutilisation en autorisant notamment le paiement d'une redevance couvrant des frais autre que ceux de la reproduction des documents. En 2013, la **révision de la directive PSI** a instauré un véritable droit à réutilisation des données du secteur public pour lequel persiste toutefois une exception pour le secteur culturel (musées, bibliothèques, archives). La loi relative à la gratuité et aux modalités de la réutilisation des informations du secteur public, dite **loi Valter de 2015**, l'a traduit en droit français en instaurant le principe de gratuité généralisé à tous les secteurs,

2.3. LOI POUR UNE RÉPUBLIQUE NUMÉRIQUE : L'OUVERTURE DES DONNÉES PAR DÉFAUT 17

toutefois, les données des services publics industriels et commerciaux restent exclues ce qui a été abrogé par la loi pour une République Numérique que nous détaillerons dans la prochaine section.

Aller plus loin : lire le guide “Les lois régulant la donnée publique” produit par Open Data France et la présentation “Quelles obligations réglementaires pour l’ouverture des données?” produite par le SGAR de la région Occitanie dans le cadre du projet Open DataLab

Enfin, dernier élément du contexte juridique, la **directive européenne Inspire du 14 mars 2007** vise à l’harmonisation et à l’ouverture des données géographiques au niveau européen. Transposée en droit français par l’ordonnance du 21 octobre 2010, elle impose aux autorités publiques de recenser les données qui rentrent dans le périmètre concerné par la directive, de créer et de maintenir les métadonnées selon les normes établies par la commission européenne, de mettre les métadonnées et les données en suivant les standards européens. Pour certaines données qui ne pourraient être ouvertes au public, la directive impose le partage entre autorités publiques.

Aller plus loin : lire le document “La directive Inspire pour les néophytes” établi par la Mission de l’information géographique (MIG) du ministère du développement durable en 2016

2.3 Loi pour une République Numérique : l’ouverture des données par défaut

Promulguée le 7 octobre 2016, la loi pour une République Numérique a fait l’objet de près de quatre ans d’annonces, de consultations et de débat. Son titre 1 porte sur l’ouverture des données publiques et rend obligatoire la mise à disposition des données publiques communicables.

La loi impose un principe d’**ouverture des données par défaut** à toutes les administrations et collectivités locales de plus de 3500 habitants et 50 agents qui concerne un **périmètre très large** comprenant les documents communiqués suite à des demandes CADA (depuis le 7 avril 2017), les “bases de données” et les données “dont la publication présente un intérêt économique, social, sanitaire ou environnemental” à partir d’octobre 2018. Rares sont les données publiques qui ne sont donc pas concernées par cette obligation d’ouverture. L’obligation concerne aussi les **entreprises délégataires d’une mission de service public** pour les données collectées ou produites à l’occasion de l’exploitation du service public faisant l’objet du contrat qui doivent être fournies gratuitement aux acteurs publics, en charge alors de les diffuser. La loi ne prévoit pas de sanctions pour les administrations qui ne respecteraient pas cette obligation.

Toutefois, la **notion de données** qui délimite le périmètre concerné par l’obligation d’ouverture ne fait pas l’objet d’une définition juridique. Classiquement, on définit les données comme une forme de représentation d’une réalité destinée généralement à être traitées par un ordinateur, à l’opposée des informations généralement destinées à être traitées par des humains. Cette définition est une parmi d’autres, la notion de données faisant l’objet de débats dans la littérature sur lequel nous n’allons pas nous étendre ici. Dans le droit, la seule référence que nous connaissons est un arrêté du 22 décembre 1981 relatif à l’enrichissement du vocabulaire informatique définit la donnée comme la « représentation d’une information sous une forme conventionnelle destinée à faciliter son traitement (en anglais : data). » L’absence de définition juridique pourrait devenir problématique dès lors que des contentieux tenteront de faire appliquer l’obligation d’ouverture des données prévus par la loi pour une République numérique.

Sur ce point, la loi pour une République Numérique consacre l’**utilisation de standards ouverts** pour la communication des documents administratifs afin de faciliter la réutilisation des données par les machines L’administration (ministères, collectivités territoriales, établissements publics...) sera dorénavant tenue, lorsqu’elle communique un document administratif au format électronique, de le mettre à disposition du citoyen “dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé”. Concrètement, cela empêche la publication de fichiers PDF là où un fichier CSV pourrait être diffusé, conformément aux principes internationaux de l’open data évoqué dans la section précédente.

Enfin, la loi **limite par décret les licences** qui peuvent être utilisées par les administrations. Le décret du 27 avril 2017 autorise deux licences : la Licence Ouverte d’Etalab et la licence OpenDataBaseLicense (nous

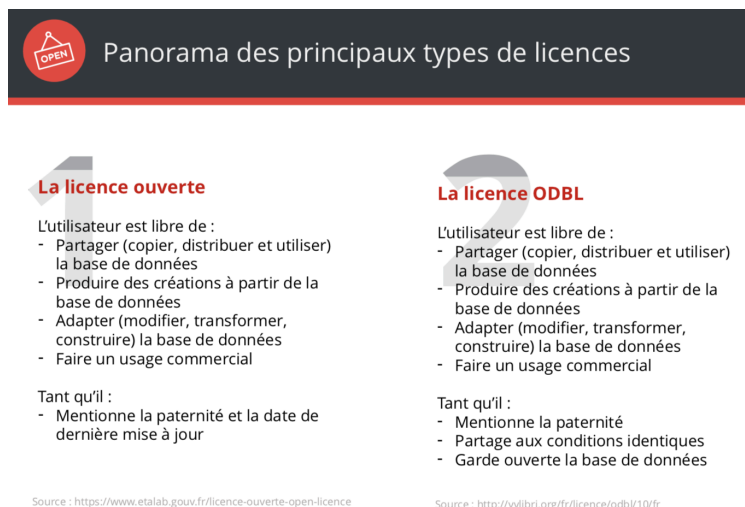


Figure 2.1: Panorama des licences de données ouvertes (source : Open Data Lab)

reviendrons à la suite sur le contenu de ces licences). Les administrations qui voudront utiliser une autre licence devront remplir une procédure d'homologation auprès des services du Premier ministre qui impose notamment un exposé des motifs ayant conduit à choisir une licence hors de la liste fixée dans le décret et une consultation des usagers affectés par la licence proposée. L'homologation doit être faite pour chaque jeu de données même si la licence reste la même.

Aller plus loin : lire l'article “On vous explique le volet « Open Data » de la loi Lemaire” publié par le site NextImpact

Ces deux licences offrent un choix aux administrations, la licence ouverte étant plus permissive que la licence ODBL qui a été conçue dans une logique de “pot commun” permettant de garantir que les données resteront ouvertes après leur réutilisation.

Aller plus loin : le guide “Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France ” de l'INRA présente dans sa fiche numéro 6 une comparaison complète de la licence ouverte et de la licence ODBL

Chapter 3

Les bénéfices attendus d'une politique d'open data pour une collectivité

Les bénéfices de l'open data sont souvent résumés à trois dimensions : transparence, innovation et modernisation. Ces trois dimensions sont juste et utiles pour repérer les bénéfices qu'on peut attendre d'une politique open data ambitieuse. Mais ces objectifs génériques peinent à rendre compte de l'utilité immédiate que l'ouverture des données peut procurer aux institutions qui s'engagent dans une démarche d'open data.

3.1 Du point de vue de l'organisation, l'ouverture des données comme un levier d'innovation ouverte

Au niveau organisationnel, l'ouverture des données peut être vue comme une manière de **valoriser le travail de l'administration**. En effet, l'ouverture des données permet de montrer au public la qualité du travail réalisé par les agents et rend visible l'effort d'objectivation des politiques publiques par la donnée. L'ouverture des données peut aussi être un moyen de présenter au public les compétences et les expertises du département. Par exemple, le département de Haute-Garonne présente les jeux de données en cinq catégories à la une du portail qui résume les compétences départementales : solidarités, déplacements, territoire, culture, jeunesse et famille.

L'open data est aussi un moyen de motiver les agents, l'ouverture des données permet de **servir l'intérêt général dans une logique de service public**. Un des arguments les plus fréquents en faveur de l'ouverture des données consiste à dire que l'ouverture des données permet de rendre au public les données qui ont été payé par l'impôt, c'est sous cet angle qu'au Royaume-Uni le *Guardian* a justifié son appel à l'ouverture des données avec sa campagne en 2006 "rendez nous les joyaux de la couronne !" En France, on peut relier les principes de l'open data aux trois grands principes du service public : - la **continuité** du service public : dès lors que les données sont ouvertes, elles peuvent continuer à être utiles pour le public même si les services qui s'appuient dessus ont été clôturés comme c'est souvent le cas avec les applications mobiles produites par les acteurs publics ; - l'**égalité** devant le service public : lorsque les données sont ouvertes, toute personne a un droit égal d'accès et de réutilisation en vertu du principe de non-discrimination des usagers au coeur de l'open data ; - l'**adaptabilité** : dès lors que les données sont ouvertes, les citoyens sont en capacité de faire évoluer les services en fonction des besoins des usagers et des évolutions techniques.

Par ailleurs, l'open data peut aussi être vu comme un **gain de temps** lorsque les agents reçoivent de nombreuses demandes de la part d'entreprises et d'acteurs associatifs pour obtenir des données. Leur ouverture permet d'avoir un point centralisé où trouver les données et éviter de renvoyer à de multiples reprises les mêmes données. Aussi, ouvrir les données brutes évite de mettre en forme les données comme c'était souvent le cas avec les fichiers PDF. Par exemple, un agent de la ville de Paris publiait régulièrement des fichiers