

Chapitre 2

Les grandes étapes de la définition des principes de l'ouverture des données

Après avoir retracé les différentes origines et filiations de l'ouverture des données, ce chapitre revient sur six moments de définition de ce qui est devenu le mouvement de l'open data. J'ai sélectionné ces épisodes, car on y voit progressivement se consolider de grands principes qui vont porter sur le processus de l'ouverture des données, les politiques publiques qui vont le définir et le cadre juridique de leur réutilisation.

Le premier épisode de cette généalogie de l'ouverture des données publiques débute en 2005 avec la rédaction de l'Open Definition par l'Open Knowledge Foundation. Fondé sur la définition de l'open source, ce texte propose des critères essentiellement juridiques qui décrivent les droits des usagers d'un savoir ouvert. Le deuxième moment que je retrace, la réunion de Sebastopol en Californie en 2007, a défini les principes de l'ouverture des données gouvernementales. Souhaitant voir leurs principes adoptés par le futur président des États-Unis, leur ambition a été satisfaite avec la signature par Barack Obama d'un mémorandum sur l'*Open Government* à son entrée à la Maison-Blanche. L'inventeur du web, Tim Berners-Lee joue le premier rôle des deux épisodes suivants. En 2009, il donnait une conférence restée célèbre lors de laquelle il réclamait l'ouverture des données brutes. En 2010, son modèle en cinq étoiles proposait une approche progressive pour que les gouvernements adoptent des standards ouverts de données. Tim Berners-Lee a ensuite conseillé le gouvernement britannique dans sa politique d'open data. Ensuite, nous retournons auprès de l'Open Knowledge Foundation lorsqu'en 2012, elle créait un outil de *benchmarking*, l'Open Data Index qui classe les États selon la publication d'une sélection de données « essentielles. » Enfin, je reviens sur l'adoption en 2013 d'une charte par les chefs d'État du G8 qui ont déclaré vouloir faire de l'ouverture des données la pratique par défaut des administrations. On le voit à travers ce résumé, ces six épisodes retracent la trajectoire d'acteurs et de projets très différents. Ils montrent comment un vocabulaire, des revendications, des normes et des pratiques ont accompagné l'élaboration des politiques d'ouverture de données.

Episode 1, "Open Definition" : des droits de

L'usager d'un savoir ouvert

En mai 2004, Rufus Pollock, chercheur en économie à l'université de Cambridge, annonçait la création de l'Open Knowledge Foundation (OKFN), une organisation à but non commercial visant à « promouvoir l'ouverture de toutes les formes de savoir [...] information, données et tous les termes synonymes¹. » Peu après la création de l'organisation, en août 2005, Pollock 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 décrire une série de conditions essentiellement juridiques permettant d'établir qu'un savoir est ouvert². La définition devait aussi servir à énumérer les licences ouvertes spécifiques au savoir et à fédérer des disciplines éparses. Pollock n'employait pas le terme « *open data* » dans son message et dans la définition, mais il signalait une prolifération de mouvements se revendiquant de l'ouverture. La définition devait servir à éviter une dilution de la philosophie de l'ouverture et à fédérer différents projets ayant pour point commun l'ouverture du savoir. Sa définition se fondait directement de l'expérience du mouvement du logiciel libre en informatique, une généalogie clairement affirmée dans le premier brouillon du texte. Pollock y créditait la définition de l'open source comme la ressource essentielle qui a servi à la rédaction de la définition, mais aussi à forger l'idée même d'ouverture.

Pour la résumer en quelques mots, l'Open Definition 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'utilisateurs 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 dite de *share alike*).

¹. OKFN, « Open Knowledge Foundation Launched », <http://blog.okfn.org/2004/05/24/open-knowledge-foundation-launched/>, consulté le 15 avril 2015.

². Open Knowledge Definition mailing list, « [okd-discuss] RFC: Open Knowledge Definition v0.1. » <https://lists.okfn.org/pipermail/okfn-discuss/2005-August/005233.html>, consulté le 15 avril 2015.

Un point a fait débat après la publication de la première version de l'Open Definition : l'exclusion des œuvres adossée à une clause non commerciale. Ce débat porte particulièrement sur le cas des licences Creative Commons. Créées en 2001 par Lawrence Lessig, un juriste de Harvard et militant de la culture libre, elles proposent une alternative au copyright en permettant aux créateurs de conserver certains droits et partager gratuitement des œuvres {Kelty:2008ue}. Chaque licence Creative Commons comporte une ou plusieurs clauses standardisées. Elles peuvent être combinées pour exiger de l'utilisateur de citer la source (*Attribution - BY*), interdire les œuvres dérivées (*Non Derivates - ND*), les usages commerciaux (*Non Commercial - NC*) et demander le partage des oeuvres dérivées avec la même licence (*Share Alike - SA*). Or, l'Open Definition stipule dans son huitième article³ que « la licence ne peut exclure l'utilisation de l'œuvre dans un domaine spécifique. Elle ne peut par exemple interdire l'utilisation de l'œuvre dans le domaine commercial. » La définition classe les œuvres publiées avec une clause non commerciale comme « fermées » au même titre qu'une œuvre diffusée selon les règles classiques du droit d'auteur. Malgré ces débats parmi les acteurs qui se revendiquent de l'ouverture, la définition du savoir ouvert a constitué une ressource précieuse pour l'ouverture des données publiques en posant la base d'un élargissement de l'*open source* au savoir. 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.

L'Open Definition a fourni une définition et des critères sans formuler de revendications ou d'exigences à l'égard des décideurs politiques. En 2007, quelques mois avant l'élection de Barack Obama, des militants de l'ouverture du savoir se sont réunis pour faire part de leurs demandes au futur président des États-Unis. Portant spécifiquement sur les données gouvernementales, ils ont réclamé une évolution radicale de leurs procédures de diffusion et des conditions de leur réutilisation.

³. Open Definition, « Définition du Savoir Libre v.1.0. » <http://www.opendefinition.org/okd/francais/>, consulté le 20 avril 2015.

Episode 2, "Sebastopol" : l'ouverture exhaustive des données primaires

Les 7 et 8 décembre 2007⁴, se tenait à Sebastopol en Californie une rencontre sponsorisée par la Sunlight Foundation, Google et Yahoo au sein des locaux de la maison d'édition O'Reilly. Cette rencontre avait pour ambition de lister dix principes de l'*open government*. Ils espéraient que les candidats à l'élection du président des États-Unis suivraient leurs recommandations. Signe des multiples facettes de l'open data, l'invitation insiste sur le fait que l'*Open Government* pourrait renouveler la participation de la société civile tout en créant des opportunités économiques pour les entrepreneurs. C'est un basculement sémantique important puisque, comme nous l'avons vu dans le chapitre précédent, l'*Open Government* désigne aux États-Unis la révélation des secrets étatiques et n'a pas de portée économique ou entrepreneuriale {Yu 2012}.



Figure 2. Les participants à la réunion à Sebastopol de l'*Open Government Data working group*. Image : David Orban sur Flickr.

Trente participants (figure 2) ont accepté l'invitation, ils ont été sélectionnés en fonction de leur affiliation à une organisation qui exige, ouvre ou réutilise des données publiques. D'autres ont été invités pour leur implication dans des projets très variés relatifs à la participation des citoyens. Parmi les participants, on peut citer

⁴. Open Government Working Group, http://public.resource.org/open_government_meeting.html, consulté le 10 avril 2015.

l'éditeur Tim O'Reilly connu notamment pour avoir popularisé l'expression « web 2.0. », Lawrence Lessig juriste et fondateur de Creative Commons, et Aaron Swartz, créateur de Reddit et militant des libertés numériques, en particulier de l'accès ouvert aux publications scientifiques. Cela vous aura peut-être sauté aux yeux, l'Assemblée était quasi exclusivement masculine à l'exception d'une seule femme.

A l'issue des deux jours de débat et d'ateliers, un communiqué final a été publié en insistant sur les bénéfices de l'*open data* pour la démocratie, l'innovation et l'amélioration du service public. Le texte a défini 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 ;
- 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 utilisateurs ;
- 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.

Maintenant que nous avons aperçu le contenu de ces deux journées et leur résultat, revenons-en au texte de l'invitation : les participants ont-ils rempli leur objectif, à savoir l'adoption de ces principes par le futur président des États-Unis ? Le 21 janvier 2009, jour de son investiture à la Maison-Blanche, Barack Obama a signé deux mémorandums sur l'*Open Government*. Après la période d'opacité des

guerres d'Afghanistan et d'Irak, ce geste devait marquer un tournant fort en faveur de la transparence. Le premier mémorandum 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. Inspiré par les principes de Sebastopol, le nouveau président se réclame de la philosophie de l'ouverture : « l'ouverture renforcera notre démocratie et promouvra l'efficacité et l'efficacite du gouvernement.⁵ » Le mémorandum a établi trois principes dont le premier sur la transparence demandait que l'information fédérale soit valorisée comme un actif stratégique pour la nation, le second encourageait les politiques de participation des citoyens à la vie publique et le troisième recommandait des méthodes innovantes et invitait à la collaboration avec les associations et les entreprises. Le président a aussi nommé Vivek Kundra, ancien *Chief Technology Officer* (CTO) de la ville de Washington comme *Chief Innovation Officer* de la Maison Blanche afin de mettre en oeuvre cette réforme. Kundra, en mars 2009, a annoncé la création de data.gov, le portail fédéral des données ouvertes, inspiré par l'expérience de Washington qu'il a pilotée. En décembre 2009, la directive *Open Government* stipulait que les administrations devaient fournir un plan d'action en vue de l'ouverture de nouvelles données en reprenant, parfois littéralement, certains des principes de Sebastopol. On retrouve encore la trace des principes de Sebastopol en 2011 dans l'*Open Government Declaration* signée par les États-Unis et sept autres pays qui a donné naissance à l'Open Government Partnership, une organisation internationale qui compte désormais plus de 70 pays membres s'engageant à faire avancer la transparence, la participation et la collaboration avec la société civile.

Cette rencontre a marqué un tournant. C'est à partir de cette date que le terme « open data » est entré dans le débat public et a donné lieu aux premières politiques d'ouverture des données. L'ouverture des données a aussi été popularisée par la mobilisation d'une figure tutélaire, l'inventeur du web Tim Berners-Lee.

⁵. Transparency and Open Government | whitehouse.gov. (s. d.). Consulté 25 mars 2020, à l'adresse <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>

Episode 3, "Raw Data Now" : l'entrée en politique 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. Le tout en moins de quinze minutes. Tim Berners-Lee⁶ s'y est présenté comme l'inventeur du web.

Dans la vidéo de la conférence dépassant aujourd'hui le million de vue, Tim Berners racontait 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 disait ressentir la même difficulté pour accéder aux données qu'à l'époque de la création du web avec les documents alors que les données déterminent une grande partie de nos vies. Il se félicitait de l'émergence des politiques d'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 estimait que l'ouverture des données implique aussi de transformer les attitudes des administrations. Il expliquait que, 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, avec son outil Gapminder, a contesté des mythes répandus sur le développement des populations dans le monde. Pour produire cet outil, Rosling a dû exiger des données à une multitude d'institutions internationales. Avec l'expression « **database hugging** », Rosling avait proposé une métaphore dans laquelle les agents 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 (figure 3).

⁶. TED, « The next web. Présenté à TED Talk. », http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html, consulté le 19 avril 2015



Figure 3. Tim Berners-Lee, lors de sa conférence TED de 2009, mimant le *database hugging*, l'attitude des administrations qui « s'accrochent » à leurs données.

Il a expliqué que 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 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 4).



Figure 4. Tim Berners-Lee appelle le public à crier « *raw data now* ».

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 mémorable : 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.

Enfin, cette conférence a eu un impact considérable dans la trajectoire de Tim Berners-Lee et dans la création de data.gov.uk. Son appel à l'ouverture des données brutes a été entendu jusqu'au 10 Downing Street. En juin 2009, lorsque le Premier ministre, Gordon Brown, a nommé Berners-Lee *information advisor* du gouvernement⁷ aux côtés de Nigel Shadbolt, un professeur d'informatique connu pour avoir fondé la « science du web » (figure 5).

⁷. BBC, « Web creator job "beyond politics." », <http://news.bbc.co.uk/2/hi/technology/8096273.stm>, consulté le 14 avril 2015.



Figure 5. Tim Berners-Lee et Nigel Shadbolt reçus par Gordon Brown au 10 Downing Street le 15 septembre 2009. Source : Number10.gov.uk⁸.

Le 21 janvier 2010, data.gov.uk a été lancé en version beta. Le 6 mai 2010, David Cameron entrait en fonction au 10 Downing Street, il avait fait de l'*open data* une des composantes essentielles de son projet de *Big Society* {Chrzanowski:2011wu}. Arrivant à la fin de leur mandat d'un an, Berners-Lee et Shadbolt sont restés conseillers du gouvernement au sein d'une nouvelle instance intitulée *Public Transparency Board* chargée de contrôler l'application du *Transparency Agenda* évoqué dans le chapitre précédent. Rufus Pollock, fondateur de l'Open Knowledge Foundation y a siégé aux côtés notamment de Tom Steinberg, un des participants de la réunion de Sebastopol et fondateur de MySociety, une organisation britannique connue pour ses projets d'innovation civique.

La conférence TED de Tim Berners-Lee a donc imposé la demande de données brutes comme une composante essentielle de l'ouverture des données. Elle a mené l'inventeur du web au cœur de la politique britannique d'ouverture des données. Un an après avoir appelé à l'ouverture des données brutes, Tim Berners-Lee a tenté de nouveau d'influencer les politiques d'open data en proposant un classement des formats de données qui incitait les administrations à lier leurs données et à les décrire par des nomenclatures partagées. Ce classement, prenant la forme d'un « modèle en cinq étoiles », a placé l'utilisation de formats ouverts comme une revendication essentielle de l'open data.

⁸. Number10.gov.uk, « PM welcomes Sir Tim Berners-Lee to Downing Street. », <http://webarchive.nationalarchives.gov.uk/20091005122636/http://www.number10.gov.uk/Page20595>, consulté le 26 juillet 2015.

Episode 4, "5-star model" : des formats ouverts et lisibles par les machines

Après avoir exigé l'ouverture des données brutes, Tim Berners-Lee a appelé à l'utilisation de formats ouverts de données. En 2010, il a proposé 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. Sur la boutique en ligne du W3C⁹, 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. (figure 6).



Figure 6. Tasse du W3C reprenant le modèle en cinq étoiles de Tim Berners-Lee.

Source : w3.org.

Dans la hiérarchie de Tim Berners-Lee, les données sont ouvertes dès la validation du premier critère. Il a considéré 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. La deuxième étoile

⁹. (http://www.cafepress.com/w3c_shop.597992118)

exigeant que les données publiées sur le web sous une licence ouverte soient lisibles par les machines et structurées. En plus des deux autres critères précédents, l'obtention de 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. Enfin, la cinquième étoile demande qu'elles soient liées à d'autres données publiées sur le web. Dans les administrations, l'utilisation de formats sémantiques, les deux derniers niveaux du modèle, réclame généralement un travail trop important de transformation des données au regard des moyens alloués aux projets d'open data.

Retenons du classement en cinq étoiles qu'il a suggéré 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. Tim Berners-Lee s'est servi de ce modèle pour composer un outil d'audit de l'ouverture des données, l'Open Data Barometer. En novembre 2009, il a créé la Web Foundation, un organisme à but non lucratif qui vise à maintenir le web libre et accessible. Elle publie depuis 2013 l'Open Data Barometer, un classement des pays selon leurs politiques d'open data qui évalue les politiques d'accès à l'information, les initiatives publiques en matière d'open data et l'ouverture d'une liste de données publiques jugées essentielles. Même si la méthodologie du baromètre est particulièrement étoffée, je vais plutôt m'intéresser à un autre classement, l'Open Data Index de l'OKFN, car il a joué un rôle déterminant en faveur de l'ouverture de certaines données. Surtout, ses critères ont consolidé la définition d'une donnée ouverte et marqué une rupture avec l'ouverture complète revendiquée dans les principes de Sebastopol en délimitant un périmètre de données essentielles, à ouvrir en priorité.

Episode 5, "Open Data Index" : un score d'ouverture et des données "essentielles"

Le 17 avril 2012, un billet sur le blog de l'Open Knowledge Foundation présentait le nouveau projet de l'organisation, l'Open Data Index. Selon Rufus Pollock, l'*open data* s'est propagé partout dans le monde, mais les données n'ont pas toujours été publiées de la bonne manière. Il explique qu'il ne suffit pas de mettre en ligne quelques feuilles de calcul sous une licence ouverte pour ouvrir les données d'une administration. L'Index porte sur une sélection de jeux de données essentiels publiés par les gouvernements et évalue leur niveau d'ouverture.

L'Open Data Index prend la forme d'un tableau de bord, qui indique en un coup d'œil le niveau d'ouverture d'une donnée. Les résultats de l'Index sont présentés dans un tableau dans lequel chaque ligne concerne un pays. Chaque cellule indique le niveau d'ouverture d'une donnée essentielle avec neuf barres. Un code couleur confirme le respect de chaque critère (vert : ouvert/rouge : fermé/bleu : incertain). Chaque jeu de données est évalué selon neuf critères très proches des principes de Sebastopol. Mis à jour chaque année, l'Index est assorti d'un classement qui valorise la progression des « bons élèves ». Sa présentation vise à identifier en un coup d'œil les pays en tête et les données les plus ouvertes (figure 7).

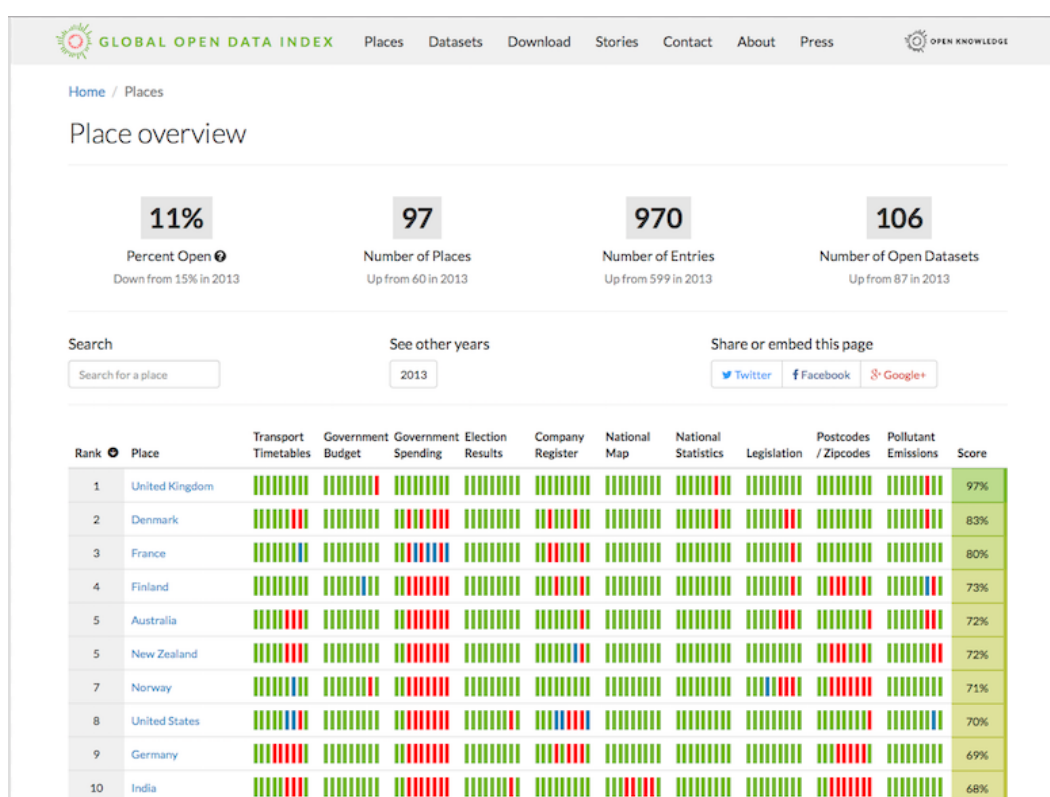


Figure 7. Résultats de l'Open Data Index 2014 présentés par pays.

L'Open Data Index repose sur la sélection de certains jeux de données devant être ouverts en priorité. De 2012 à 2014, ces jeux de données étaient définis par les intitulés suivants :

- les résultats des élections nationales au niveau de la circonscription ;
- les registres des entreprises (sans les informations financières sur le bilan notamment) ;

- les données géographiques sur le territoire national à l'échelle 1:250000 ;
- les dépenses de l'État au niveau transactionnel ;
- le budget de l'État ;
- la législation (lois et décrets) ;
- les données statistiques nationales sur l'économie et la démographie ;
- la base de données des codes postaux avec leur géolocalisation ;
- les horaires des transports publics nationaux ;
- les données environnementales sur les principaux polluants.

Par rapport à l'Open Definition, aux principes de Sebastopol et aux travaux de Tim Berners-Lee, l'Index a marqué plusieurs ruptures importantes. Plutôt que de réclamer l'ouverture complète des données sans évaluer préalablement leurs conditions concrètes de production et de réutilisation, il a délimité une sélection de données essentielles à ouvrir prioritairement et partout dans le monde. L'OKFN estimait que ces données sont produites dans chaque pays, essentielles à la transparence et à la création de services aux citoyens.

À travers ces cinq épisodes, nous avons abordé des initiatives locales et des instruments dont on a du mal à mesurer la portée concrète au niveau international. Mais en 2013, l'open data est apparu à l'agenda des discussions des chefs d'État participants à la réunion du G8 en Irlande du Nord. Comment la charte qui en résulte a-t-elle traduit ces initiatives dans le langage de la diplomatie et des engagements internationaux des États ?

Episode 6, "G8" : la reconnaissance de données à forte valeur

Les 17 et 18 juin 2013 à Lough Erne en Irlande du Nord, le Premier ministre britannique, David Cameron, accueillait la réunion du G8, la rencontre des huit chefs d'État parmi les plus grandes puissances économiques mondiales (Allemagne, Canada, États-Unis d'Amérique, France, Royaume-Uni, Italie, Japon, Russie). Les journalistes en ont essentiellement retenu les déclarations autour de la Syrie et de la

lutte contre l'évasion fiscale. David Cameron entendait pourtant faire de Lough-erne le « sommet de la transparence. » L'agenda comportait une session (figure 8) sur la publication d'information sur les industries extractives, la transparence de la propriété des terres et l'adoption d'une charte sur l'*open data*.



Figure 8. Une session de travail des chefs d'État lors du G8 de 2013¹⁰.

La charte part du constat que l'ouverture des données publiques pourra renouveler la croissance économique et rendre les gouvernements plus redevables et efficaces. Le préambule détaille les bénéfices de l'ouverture des données : 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... Pour éviter que l'ouverture des données publiques ne soit une « opportunité manquée », les chefs d'État du G8 ont décidé de l'adoption de cinq principes pour régir l'accès aux données.

Les trois premiers principes établissent les conditions d'ouverture des données puis les deux derniers fixent deux objectifs : l'amélioration de la gouvernance et le soutien à l'innovation. Le premier point de la charte annonce que les pays signataires s'engagent à faire de l'ouverture la nouvelle norme pour les données publiques. Chaque État membre devra publier un plan d'action pour mettre en œuvre la charte du G8. Dans son deuxième principe, la charte promet la publication de données de qualité. Les gouvernements s'engagent à publier les données dès que possible, « sous leur forme originale et non modifiée, et au plus fin niveau de granularité disponible ». Cette dernière ressemble fortement aux exigences de

¹⁰. G8UK sur FlickrR, <https://www.flickr.com/photos/g8uk/>, consulté le 1 août 2016.

données « primaires » des principes de Sebastopol ou de la demande de données brutes de Tim Berners-Lee. Selon le troisième principe, les données doivent être publiées dans des portails uniques par pays qui n'exigent pas l'enregistrement des utilisateurs. Elles doivent aussi être gratuites et « dans des formats ouverts chaque fois que possible. » Dans le quatrième principe, les États du G8 s'engagent à partager leur expertise technique avec les pays du monde entier, notamment au sein d'initiatives multilatérales telles que l'Open Government Partnership. Ils déclarent vouloir identifier les jeux de données à ouvrir en priorité avec les organisations de la société civile. Dans le cinquième principe, les gouvernements s'engagent à développer la culture de l'ouverture des données et à encourager les organisations de la société civile à se saisir des données.

À travers ce résumé, on voit donc que la charte du G8 s'est inscrite dans la continuité des définitions de l'*open data évoquées précédemment. En particulier, elle a repris la majeure partie des principes de l'*Open Government Data* établis à Sebastopol. La charte s'est inspirée de l'Open Data Index en sélectionnant des données à forte valeur ajoutée devant être ouvertes en priorité. La charte indique que l'ouverture des données devra se faire de manière progressive en se concentrant d'abord sur les données « à forte valeur ajoutée », celles pour lesquelles une multitude d'usages potentiels est projetée.

Le suivi de la charte sur l'open data ne figure plus au programme des sommets. Le G7 de juin 2014 était dédié à la crise en Crimée et celui de 2015 en Allemagne n'a pas évoqué la charte, le pays hôte n'ayant pas publié son plan d'action. Bien qu'elle ne soit plus au programme des débats du G7, la 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 Unies¹¹.

Adopté par une quinzaine de pays dans le monde dont la France, la Charte internationale de l'open data fixe cinq grands principes pour l'ouverture des données :

- faire de l'open data la pratique par défaut des administrations ;
- publier les données en temps opportun (*timely*), désagrégées et de qualité ;

¹¹.

<https://opendatacharter.net/seventeen-governments-adopt-the-new-international-open-data-charter/>

- publier les données sur des portails centraux, sans enregistrement préalable, avec une licence ouverte ;
- fournir les données dans des formats structurés et normalisés favorisant l'interopérabilité et la réutilisation
- ouvrir les données pour améliorer la gouvernance et encourager la participation citoyenne
- favoriser le développement inclusif et l'innovation en donnant du pouvoir d'agir à tous.

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. Surtout, elle a donné lieu à la création d'une organisation (l'Open Data Charter) en charge de promouvoir ces principes et de définir une stratégie pour faire aboutir ses objectifs.

Conclusion

Lors de ces six épisodes, nous avons fait la navette essentiellement entre les États-Unis et la Grande-Bretagne à la rencontre d'acteurs, très majoritairement masculins, aux profils très divers (des informaticiens mais aussi des juristes, des experts de l'innovation, des diplomates ou encore des gouvernants. Même si les intitulés des professions varient, on remarque que les acteurs qui ont défini les grands principes de l'open data étaient d'abord portés sur les questions techniques de la mise à disposition des données et moins sur les conditions politiques ou organisationnelles de l'ouverture. C'est particulièrement vrai au sujet des principes issus de la rencontre de Sebastopol qui, encore aujourd'hui, définissent ce qu'on entend classiquement par « open data. » Yu et Robinson {Yu:2012wv} ont montré que les principes de Sebastopol étaient guidés essentiellement par des considérations techniques. En effet, l'ensemble des principes portent sur la manière dont les données doivent être mises à disposition mais ne s'intéressent pas au contenu des données. Les auteurs expliquent qu'un régime pourrait ainsi ouvrir des données en respectant ces principes sans que cela contribue à renforcer la transparence et la réédition des comptes, promus comme une des bénéfices de l'ouverture de données. Par exemple, le gouvernement de Singapour, où la liberté d'expression est restreinte et le même parti au pouvoir depuis 1959, a investi massivement dans l'ouverture des données publiques. Les données publiées sur data.gov.sg ont été sélectionnées pour la capacité à créer de la valeur et des

services pratiques sans renforcer la transparence du régime, le portail servant à générer une impression de transparence et d'ouverture {Stevens:2019io}.

On peut d'autre part souligner une tension importante entre deux modèles, l'un qui prône une ouverture exhaustive, l'autre qui insiste sur la mise à disposition prioritaire de données jugées essentielles. D'un côté, les principes de Sebastopol proposent un modèle fondé sur l'ouverture exhaustive de toutes les données publiques sous leur forme primaire qu'on retrouve aussi dans l'appel de Tim Berners-Lee à la publication de l'ensemble des données brutes. De l'autre côté, l'Open Data Index et la charte internationale de l'open data ont introduit une sélection de données essentielles qui doivent être ouvertes en priorité en fonction de la demande et du potentiel de la réutilisation. Le premier modèle propose des critères de l'ouverture qui visent à réduire les frictions de la réutilisation des données par les machines. Ces critères sont à la fois juridiques en réclamant l'utilisation de licences standardisées et techniques en demandant l'ouverture des données dans des standards aux spécifications ouvertes et lisibles par les machines. Ces revendications s'inscrivent dans la lignée des mouvements du logiciel libre qui réclament ce que Kelty (2008) qualifie de « modifiabilité », la capacité de transformer un objet, de le réutiliser dans un contexte différent ou encore de l'améliorer. On retrouve essentiellement ces critères dans l'Open Data Index, mais ce deuxième modèle apporte une nouvelle dimension puisqu'il porte sur le contenu même des données. La charte du G8 puis la charte internationale de l'open data se situent en quelque sorte sur une ligne de partage en demandant à la fois l'ouverture de l'ensemble des données par défaut dès leur production et en désignant des données essentielles à ouvrir en priorité.

Début 2018, l'Open Data Charter, l'organisation en charge de la promotion de la charte internationale de l'open data a publié sa nouvelle stratégie de développement. Passé relativement aperçu, ce document dénote pourtant d'un tournant majeur dans les principes de l'open data. L'organisation y acte des limites du modèle fondé sur l'ouverture exhaustive des données qui a poussé les administrations à publier des données en masse sans forcément se préoccuper de leur utilisabilité. La stratégie incite à un basculement vers un modèle nommé « *publishing with a purpose* » (publier avec un objectif) dans lequel les administrations travaillent avec des communautés d'utilisateurs de données pour résoudre des problèmes publics, s'engagent à publier des jeux de données essentiels pour relever ce défi et à apprendre puis adapter pour maximiser l'impact de l'ouverture des données.

Il est de plus en plus reconnu que l'ouverture des données de manière isolée est moins efficace qu'elle ne peut l'être si elle vise à résoudre des problèmes

politiques spécifiques - que "publier avec un objectif" peut apporter plus que "publier et ils viendront". [...] Nous pensons que la meilleure façon d'inciter les gouvernements à devenir "ouverts par défaut" est d'encourager les petits pas qui donnent des résultats rapides. Plutôt que de plaider pour que les gouvernements ouvrent autant de données que possible, aussi vite que possible, nous élaborons des conseils pratiques pour que les réformateurs gouvernementaux ouvrent des jeux de données de la manière la plus susceptible de produire des bénéfices spécifiques et tangibles pour les citoyens¹².

Ces enjeux liés à l'utilisabilité et à l'impact des données seront au coeur de la deuxième partie de cet ouvrage dans laquelle je dresserai un bilan critique de l'ouverture des données publiques en France au regard des promesses de transparence, d'innovation et de modernisation de l'État formulées autour des politiques d'open data. Mais, avant cela, il nous faut comprendre comment les principes de l'open data ont été diffusés et mis en œuvre en France.

¹². Publishing with Purpose—Open Data Charter—Medium. (s. d.). Consulté 25 mars 2020, à l'adresse <https://medium.com/@opendatacharter/publishing-with-purpose-introducing-our-2018-strategy-ddbf7ab46098>