

Une petite histoire d'Etalab : Comment l'open data s'est institutionnalisé en France



Samuel GOËTA

Co-fondateur de l'entreprise coopérative Dataactivist

« Etalab » est le nom de la structure administrative qui a été créée en France en 2011 pour promouvoir l'ouverture des données publiques. Initialement tournée vers le développement de services privés innovants utilisant ces données, cette structure s'est donné ensuite d'autres objectifs : créer une communauté d'utilisateurs, transformer l'action de l'État. Parallèlement, ses critères de réussite ont évolué : au début, il s'agissait de maximiser le nombre des fichiers rendus accessibles, progressivement l'accent a été mis sur des critères de qualité, pour aboutir à la notion de « données de référence » présente dans la loi de 2016.

Les 7 et 8 décembre 2007, c'est lors d'une rencontre dans les locaux de l'éditeur O'Reilly à Sebastopol en Californie que 30 militants des libertés numériques, entrepreneurs et chercheurs ont défini ensemble huit grands principes de l'ouverture des données qui depuis définissent classiquement ce qui constitue une donnée ouverte¹. Dix ans plus tard, la France est aujourd'hui le premier pays à avoir inscrit dans la loi un principe d'open data par défaut avec la loi pour une République numérique de 2016 qui instaure une obligation d'ouverture des données pour les institutions, les collectivités locales de plus de 3500 habitants et 50 agents ainsi que tous les acteurs investis d'une mission de service public.

Mais comment ces grands principes de l'open data ont-ils été importés en France ? Quels en ont été les passeurs ? Pour répondre à ces questions, je vais revenir sur la trajectoire de la mission Etalab qui, depuis 2011, au sein de l'Etat a été en charge de coordonner l'ouverture des données publiques en France². Cet article retrace à très grands traits la trajectoire d'Etalab, une institution qui doit être replacée dans une histoire plus longue, celle de la longue émergence des droits d'accès et de réutilisation de l'accès et la réutilisation des informations publiques³. Retenons que c'est essentiellement par l'Union européenne que le droit à la réutilisation, c'est-à-dire à l'exploitation de données obtenues par le droit d'accès à l'information publique (loi CADA de 1978 en France) notamment pour la création de services, s'est imposé en France. Dans les années 2000, la Commission européenne a multiplié les études sur le potentiel économique de la réutilisation de l'information publique, évaluant jusqu'à 200 milliards d'euros par an la valeur de leur circulation optimale dans les pays de l'Union⁴ et a adopté plusieurs directives à

1. Yu H. & Robinson D.G. (2012), The New Ambiguity of "Open Government", *UCLA Law Review* 178, pp. 178-208.

2. Cet article résume le second chapitre de la thèse de doctorat en sociologie de l'auteur.

Goëta S. (2016), *Instaurer des données, Instaurer des publics : une enquête sociologique dans les coulisses de l'open data*, Télécom ParisTech, Paris.

3. Ronai M. (1997), Données publiques : accès, diffusion, commercialisation, *Problèmes politiques et sociaux* 773-774, p. 68 ; Boustany J. (2013), Accès et réutilisation des données publiques. Etat des lieux en France, *Les cahiers du numérique* 9(1), 21-37 ; Trojette M. A. (2013), *Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes ?* Rapport au Premier Ministre.

4. Vickery G. (2011), *Review of recent studies on PSI re-use and related market*, Commission européenne.

ce sujet dont la directive Public Sector Information (PSI) de 2003 qui a fixé des règles minimales pour faciliter le droit à la réutilisation des données.

Plus de sept ans après sa création, Etalab reste une structure administrative récente aux attaches politiques et administratives instables. Nous verrons au fur et à mesure des trois grandes phases qui découpent cet article que l'action, les priorités et les équipes d'Etalab ne cessent d'évoluer témoignant de l'ancrage mouvant de l'ouverture des données dans les pratiques de l'administration.

De l'APIE à Etalab, un demi-tour en faveur de la gratuité des données

En octobre 2008, le gouvernement présentait un plan de développement de l'économie numérique pour le quinquennat, intitulé « France Numérique 2012 » qui proposait, dans son action 39, « la création d'un portail unique d'accès aux données publiques » dont l'Agence pour le Patrimoine Immatériel de l'État (APIE) devait piloter la conception. Cette agence avait été créée à la suite du rapport Lévy-Jouyet de décembre 2006 pour générer des revenus issus du « patrimoine immatériel de l'État » dont les données publiques faisaient partie. La création de ce portail s'inscrivait aussi dans le cadre de la transposition de la directive PSI de 2003. Dans le rapport dirigé par le député Franck Riester en 2010, le nom « État lab » a été proposé pour dénommer le portail que devait concevoir l'APIE pour « développer des services innovants à partir de données publiques ». Il n'est pas question alors de transparence ou de reddition des comptes aux citoyens.

Malgré ce rapport qui plaçait l'APIE au cœur du dispositif, l'agence est progressivement dessaisie du dossier au profit d'Etalab, une nouvelle structure dédiée à la diffusion gratuite des données publiques. Selon plusieurs personnes interrogées, un voyage d'étude à Washington en 2010 auquel participaient les conseillers numériques de l'exécutif ainsi que la secrétaire d'État en charge des questions numériques aurait convaincu les membres du gouvernement de demander le renvoi de l'APIE et de redéfinir le projet en faveur de la gratuité des données. D'autre part, plusieurs acteurs interrogés ont interprété l'accélération du développement de data.gouv.fr comme une volonté pour la majorité de rattraper son « retard » face aux collectivités locales d'opposition ayant engagé une politique d'open data (Rennes, Nantes, Montpellier, Paris, région PACA...). Enfin, de manière plus officieuse, Etalab aurait été créée dans l'optique de la campagne présidentielle pour faire apparaître Nicolas Sarkozy comme un candidat « transparent ». La création de la mission Etalab s'est en effet précipitée quelques mois avant le début de la campagne officielle.

Le 30 juin 2010, le conseil de modernisation des politiques publiques, un organe interministériel en charge de la révision générale des politiques publiques (RGPP), a décidé de la création d'un « État lab », un « portail Internet recensant les données existantes et permettant leur réutilisation ». Le conseil des ministres du 24 novembre 2010 a annoncé la mise en ligne de ce portail avant la fin de l'année 2011 et a officiellement dessaisi l'APIE du dossier. « État lab » ne désigne plus le portail, devenu data.gouv.fr, mais la mission en charge de sa création. Sa direction a été attribuée à Séverin Naudet, ancien vice-président du site de partage de vidéos Dailymotion, qui a été nommé en 2007 « conseiller spécial sur Internet et le multimédia » de François Fillon. Le 21 février 2011, le décret 2011-194 a créé la mission « Etalab » placée sous l'autorité du Premier ministre et rattachée au secrétaire général du Gouvernement. Le 17 octobre 2011, Etalab a publié la Licence Ouverte qui acte de la gratuité des données publiées, autorise les usages commerciaux et impose aux réutilisateurs de citer la source.

En juin 2011, un prototype de portail a été lancé, mais data.gouv.fr n'était alors qu'une coquille vide sans les données des administrations. Le 26 mai 2011, le Premier ministre, François Fillon,

a publié un décret⁵ et une circulaire⁶ adressés aux ministres, secrétaires d'État et préfets. Le décret imposait que la liste des données publiques soumises à redevance soit arrêtée par décret afin de faire de la gratuité la norme à partir du 1^{er} juillet 2012. Désignant un interlocuteur unique pour Etalab dans chaque ministère, la circulaire de mai 2011 exigeait, dans un délai d'un mois, une rencontre avec ses correspondants et dépasser le nombre de jeux de données publiés sur data.gov aux États-Unis et data.gov.uk au Royaume-Uni.



Figure 1 : Page d'accueil de data.gouv.fr lors de son lancement en décembre 2011

La première version de data.gouv.fr (figure 1) a été lancée le 5 décembre 2011, conformément au calendrier du projet. Etalab a particulièrement communiqué sur les 352 000 jeux de données publiés excédant les chiffres annoncés pour data.gov aux États-Unis et data.gov.uk au Royaume-Uni. Cette obsession du chiffre a conduit à découper les jeux de données de l'INSEE en un fichier par commune pour une même base afin de « gonfler » le nombre de jeux de données publiés. À la suite de l'élection présidentielle et de l'alternance à l'Élysée, plusieurs personnes interrogées ont douté de l'avenir de la mission Etalab et du maintien en ligne de data.gouv.fr. L'incertitude sur la politique d'open data a été exacerbée par un article publié dans Les Échos en octobre 2012 qui annonçait que certaines administrations pourraient commercialiser leurs données publiques⁷. Les membres du gouvernement avaient pourtant signé une charte de déontologie qui disposait que le gouvernement « mène une action déterminée pour la mise à disposition gratuite et commode sur Internet d'un grand nombre de données publiques⁸. » Cette incertitude rappelle que l'ouverture des données a reposé en grande partie sur les attaches politiques de la mission Etalab. En ouvrant des données, les agents ont souvent répondu à une injonction politique extraordinaire avant de participer à une pratique normale de l'administration.

5. Premier ministre. Décret n° 2011-57 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'État et ses établissements publics administratifs. Journal Officiel de la République Française. 27 mai 2011.
6. Premier ministre, Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'État « data.gouv.fr » par la mission « Etalab » et l'application du droit de réutilisation des informations publiques. Journal Officiel de la République Française. 27 mai 2011.
7. Les Echos, « Open Data : l'État pourrait renoncer à la gratuité de certaines données publiques », http://www.lesechos.fr/journal20121017/lec2_high_tech_et_medias/0202329690871-open-data-l-etat-pourrait-renoncer-a-la-gratuite-des-donnees-publiques-501147.php, consulté le 15 décembre 2014.
8. Numérama, « Internet et l'Open Data dans la déontologie du gouvernement Ayrault », <http://numerama.com/magazine/22534-internet-et-l-open-data-dans-la-deontologie-du-gouvernement-ayrault.html>, consulté le 15 décembre 2014.

Vers une administration générale des données

En décembre 2012, les services du Premier ministre ont annoncé la nomination d'Henri Verdier à la tête d'Etalab. Ancien directeur de Cap Digital, pôle de compétitivité des entreprises du secteur numérique en Ile-de-France, il a aussi fondé MFG Labs, une entreprise spécialisée dans l'exploitation de données massives. Au cours de l'année 2013, de nouvelles orientations ont été données à la mission Etalab qui a changé de rattachement pour rejoindre le SGMAP (Secrétariat Général à la Modernisation de l'Action Publique). Ce changement de rattachement a eu des conséquences sur le discours des agents d'Etalab. Henri Verdier a présenté régulièrement l'ouverture des données publiques comme une manière de transformer l'action de l'État avant d'insister sur son potentiel économique comme le faisait son prédécesseur. La feuille de route du gouvernement en matière d'open data a été publiée en février 2013⁹ pour donner de nouvelles orientations à la mission Etalab. Le premier changement portait sur les objectifs assignés à la mission. Plutôt que d'être évaluée sur la quantité de données publiées selon un objectif d'exhaustivité proche des préconisations des principes de Sebastopol, Etalab doit désormais publier des jeux de données « stratégiques » qui seront identifiés lors de six débats thématiques avec la société civile. A cette fin, l'équipe d'Etalab s'est assurée que la France allait améliorer sa place dans le classement international de l'ouverture des données essentielles de l'Open Knowledge Foundation, l'Open Data Index. Après avoir obtenu une 16^e place en 2013, plusieurs données « essentielles » (la base de données des lois, décrets et ordonnances et les codes postaux) ce qui a hissé la France à la 3^e place du classement. Le gouvernement a alors engagé une campagne de communication autour de ce résultat (figure 2).



Figure 2 : Illustration diffusée sur le compte Twitter officiel du gouvernement français suite à la publication des résultats de l'Open Data Index.

En outre, la feuille de route annonçait le lancement d'une nouvelle version du portail data.gouv.fr. Pour sa refonte, Henri Verdier déclarait s'inspirer des principes proposés dans le livre qu'il a coécrit avec Nicolas Colin, *L'Âge de la multitude*¹⁰, dans lequel ils développent l'idée selon laquelle la richesse dans l'économie numérique dépend de capacité d'un acteur à capter la

9. Etalab, « La feuille de route du Gouvernement en matière d'ouverture et de partage des données publiques », <http://www.etalab.gouv.fr/article-la-feuille-de-route-du-gouvernement-en-matiere-d-ouverture-et-de-partage-des-donnees-publiques-115767801.html>, consulté le 12 décembre 2014.

10. Colin N. & Verdier H. (2012), *L'Âge de la Multitude. Entreprendre et gouverner après la révolution numérique*, Paris, Armand Colin.

valeur de la « multitude. » Le portail a été remis à plat à l'issue d'une consultation lors de laquelle « l'écosystème » était invité à répondre à un questionnaire en ligne et à participer à des ateliers de « codesign. » Présenté comme un espace de mise en relation entre l'administration et les réutilisateurs, le nouveau portail permet de republier un fichier après l'avoir traité afin que l'administration puisse bénéficier du travail de la « multitude ». Lancé le 18 décembre 2013 à Matignon par le Premier ministre, le nouveau data.gouv.fr était présenté comme une plateforme qui, selon les éléments de langage, « permet de "faire vivre" les données et de rencontrer des innovateurs permettant de faire naître de nouveaux services. » Dans sa communication, Etalab insistait tout autant sur la valeur des données publiées sur data.gouv.fr que sur celle de la « communauté » qui s'y active. Cela s'est traduit dans les objectifs opérationnels assignés à la mission Etalab qui doit accroître le nombre d'utilisateurs actifs de la plateforme et de réutilisations recensées sur data.gouv.fr sur son site, comme on peut le voir dans le projet de loi de finances 2015 (figure 3). Avec la refonte du portail, Etalab s'est donc vue attribuer de nouveaux objectifs qui ne consistent pas uniquement à ouvrir les données, mais à les « faire vivre » en s'assurant de leur réutilisation et de l'existence d'une communauté autour des données.

INDICATEUR 7.2 : Ouverture et diffusion des données publiques							
(du point de vue du citoyen)	Unité	2012 Réalisation	2013 Réalisation	2014 Prévision PAP 2014	2014 Prévision actualisée	2015 Prévision	2017 Cible
Nombre de ressources en open data (site « data.gouv.fr »)	Nombre	SO	SO	SO	36 000	37 000	40 000
Nombre de contributeurs actifs (site « data.gouv.fr »)	Nb de comptes actifs	SO	SO	SO	3 500	4 000	10 000
Nombre de réutilisations (site « data.gouv.fr »)	Nb	SO	SO	SO	1 400	2 000	5 000

Figure 3 : Indicateurs de performances de l'ouverture des données publiques dans le projet de loi de finances 2015.

Du point de vue des données, la mutation la plus importante d'Etalab concerne la création d'une fonction d'Administrateur Général des Données (AGD), traduction française du rôle de Chief Data Officer attribué dans plusieurs administrations états-uniennes. Henri Verdier a été nommé à cette fonction par décret le 19 septembre 2014. L'AGD a pour mission de coordonner l'action des administrations en matière d'« inventaire, de gouvernance, de production, de circulation et d'exploitation des données. » Dans le cadre de ses missions, il peut être saisi par tout citoyen ou toute personne morale et doit remettre chaque année un rapport au Premier ministre sur la gouvernance des données. L'équipe de l'AGD¹¹ a conduit plusieurs expérimentations avec des administrations notamment sur des domaines variés tels que la consommation électrique de l'État, les vols de voiture ou l'emploi. Henri Verdier a ensuite été nommé en septembre 2015 directeur interministériel du numérique et du système d'information et de communication de

11. Administrateur Général des Données : « L'équipe », <https://agd.data.gouv.fr/lequipe/>, consulté en juin 2016.

l'Etat (DINSIC). Laure Lucchesi, spécialiste de la transformation numérique et des stratégies fondées sur la donnée, lui a succédé à la direction d'Etalab.

De l'ouverture des données à l'ouverture des gouvernements

A partir de 2014, une autre évolution importante pour Etalab a été l'entrée de la France dans le Partenariat pour le Gouvernement Ouvert (Open Government Partnership). En avril 2014, lors de la conférence de Paris sur le gouvernement ouvert, Henri Verdier a annoncé que la France allait rejoindre cette organisation internationale créée en 2009 par le président Obama, pour promouvoir les « bonnes pratiques » en matière de transparence des États, la participation des citoyens et la collaboration avec la société civile. En plus de leurs missions liées à l'ouverture des données, les agents d'Etalab assurent une grande partie de la représentation de la France au sein de cette institution internationale et élaborent, en concertation avec la société civile, les engagements que prend le gouvernement dans le cadre de son plan d'action. Ces nouvelles missions liées à des enjeux de participation sont critiquées en particulier par l'association Regards Citoyens qui considère qu'elles détournent Etalab de sa mission : « il est temps qu'Etalab se recentre sur les actions concrètes simples et rapides au cœur de ses missions pour vraiment faire avancer la transparence et l'Open Data. »¹² Par la suite, la France a intensifié son engagement dans le Partenariat en rejoignant son comité directeur en août 2014 puis en prenant sa présidence en 2016. En décembre 2016, Paris a accueilli le sommet annuel du Partenariat pour un Gouvernement Ouvert, un événement coordonné par Etalab et accueillant près de 4 000 participants de 140 nationalités venus débattre de transparence de l'action publique, la participation citoyenne et l'innovation démocratique. Conçu comme une « COP21 de la démocratie », l'évènement a connu médiatiquement un certain « désintérêt » selon un journaliste de Mediapart¹³ et un collectif d'associations de défense des libertés numériques, dont la Quadrature du net, la Ligue des droits de l'homme, l'April ou Framasoft, a annoncé qu'il ne participerait pas à ce sommet dénonçant un « leurre » du fait de la surveillance des citoyens en état d'urgence et de l'opacité de certaines décisions gouvernementales¹⁴.

Promulguée le 7 octobre 2016, après près de quatre ans d'annonces, de consultations et de débat, la loi pour une République Numérique impose un principe d'ouverture des données par défaut à toutes les administrations et collectivités locales de plus de 3500 habitants et 50 agents. La loi crée une nouvelle mission pour Etalab dans son article 14, celle de coordonner le service public de la donnée dont la mission est de garantir la qualité et la disponibilité de « jeux de données de référence » qui présentent le plus fort impact économique et social comme la base SIRENE des entreprises (ouverte sur data.gouv.fr depuis janvier 2017), la base adresses ou encore le cadastre (ouvert depuis octobre 2017). Cette nouvelle mission rompt en quelques sortes avec l'obsession de la quantité de données des premiers temps d'Etalab pour garantir la qualité de certaines données très réutilisées.

La trajectoire d'Etalab au cours de ces sept dernières années que nous venons de parcourir souligne l'instabilité et la fragilité de la politique publique d'ouverture des données. En effet, comme on a pu le voir à travers l'alternance, l'évolution de son contexte légal et l'attribution régulière de nouveaux objectifs en matière d'open government ou de réutilisation des données, cette structure reste très liée à ses attaches politiques et à un contexte législatif mouvant. Par exemple, depuis novembre 2017, Etalab a encore changé de rattachement quittant le Secrétariat Général de la Modernisation de l'Action Publique (SGMAP) pour rejoindre la Direction

12. Regards Citoyens, « La France presidera-t-elle l'Open Communication Partnership? », <https://www.regardscitoyens.org/la-france-presidera-t-elle-oupen-communication-partnership/>, consulté en juin 2015.

13. Mediapart, « Henri Verdier : « Vers un développement démocratique durable? » », <https://www.mediapart.fr/journal/culture-idees/061216/henri-verdier-vers-un-developpement-democratique-durable?onglet=full>, consulté en décembre 2017.

14. NextInpact, « À l'approche du sommet mondial de l'OGP, les consultations en ligne dans la tourmente », <https://www.nextinpact.com/news/102389-a-l-approche-sommet-mondial-ogp-consultations-en-ligne-dans-tourmente.htm>, consulté en décembre 2017.

interministérielle du numérique et du système d'information et de communication de l'Etat (DINSIC).

Conclusion

On peut retenir de la trajectoire d'Etalab que cette structure a contribué à porter l'attention sur les données publiques, bien au-delà de la question de leur ouverture. Le rapport annuel de 2015 de l'AGD l'explique bien en invitant à s'intéresser aux « données de gestion » produites dans les systèmes d'information de l'État : « La plupart des données existantes sont aujourd'hui produites dans de grands systèmes de gestion informatisés, et ne sont pas connues ni repérées comme telles. Une histoire connue dans les communautés open data concerne cette grande municipalité qui souhaitait ouvrir son portail d'open data et recherchait dans ce but des données concernant les pratiques culturelles. Il lui fallut près d'un an pour réaliser que l'application de gestion des bibliothèques municipales recelait un trésor [...] De telles données, issues des grands systèmes de gestion, représentent aujourd'hui un sujet central de la gouvernance de la donnée » Cette description des données de gestion correspond, à peu de choses près, aux « sources administratives » de la statistique publique évoquées par Desrosières¹⁵ lorsqu'il distingue deux sources de la statistique publique : d'une part, les « enquêtes » produites spécifiquement par des institutions dédiées selon des normes scientifiques et d'autre part, les « sources administratives » issues de services « dont les activités de gestion impliquent la tenue, selon des règles générales, de fichiers ou de registres individuels, dont l'agrégation n'est qu'un sous-produit, alors que les informations individuelles en sont l'élément important, notamment pour les individus ou les entreprises concernés. » Si les sources administratives sont utilisées depuis longtemps par les statisticiens, les politiques d'open data, dont la mission Etalab a été la cheville ouvrière depuis près de sept ans, ont contribué à mettre en lumière les données brutes de l'Etat comme le matériau informationnel de la transformation et de l'ouverture du gouvernement.

15. Desrosières A. (2005), Décrire l'Etat ou explorer la société : les deux sources de la statistique publique, *Genèse* 58, 4-27.