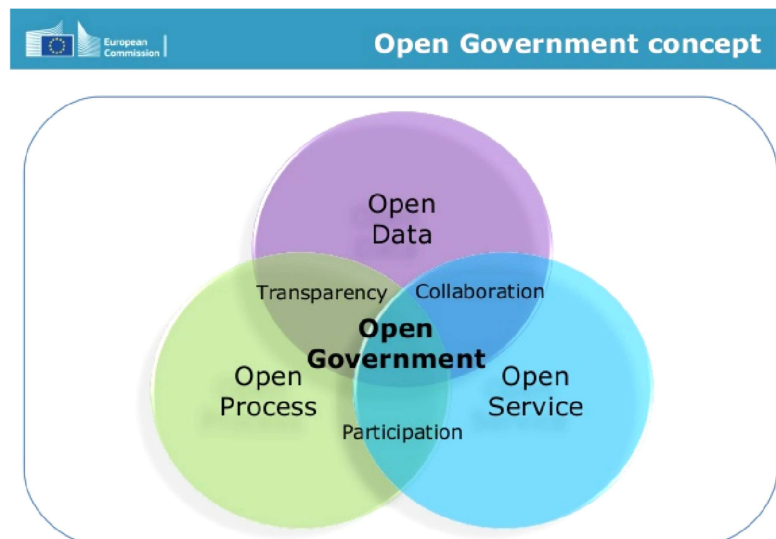


# Open data et eGovernment : au delà des idées reçues, quelques bonnes pratiques

Posted on 2014-03-26 by [Isabelle Boydens](#)

Né aux États-Unis, le mouvement *open data* est officialisé depuis la loi « *Freedom of Information Act* » de 1966. Son ampleur s'est accrue à la fin des années 1990 avec l'émergence d'Internet. L'encyclopédie libre [Wikipedia](#) fondée en 2001 constitue un exemple bien connu d'information ouverte au grand public et alimentée par celui-ci. Son contenu, dont la qualité est nécessairement inégale, évolue au gré du concours des contributeurs censés respecter un ensemble de bonnes pratiques.



La notion d'*open data* s'applique aussi à l'administration publique. La Commission européenne en fournit la définition suivante : « *Open data refers to the idea that certain data should be freely available for use and re-use* ». Comme l'illustre le schéma ci-contre, les *open data* constituent un sous-ensemble de l'*Open Government* dont l'objet consiste en la mise en ligne de processus et services transversaux au sein de l'administration mais aussi entre celle-ci

et ses partenaires. Dans ce contexte, les *open data* ont pour objectif général de produire de la valeur en vue :

- de la création de nouveaux services générateurs de gains via le partage collaboratif de données;
- de favoriser l'innovation (à travers l'émergence d'usages nouveaux et inattendus) ;
- d'encourager la participation active et spontanée des citoyens dans la vie sociale et politique et d'accroître la transparence du gouvernement.

Un état de l'art de la question a été récemment présenté à Gand en février 2014 lors des *Data Days*. Voici un exempl



de *open data* très classique, dans le cadre des « *smart cities* » : celles-ci aident à localiser dans la ville de Gand les points d'accès pour les personnes handicapées circulant en chaise roulante. En Belgique, on trouve une recension d'*open data* sur le site fédéral géré par l'ASA/DAV : Public Sector Information. La question est en ce moment d'actualité sur le plan réglementaire européen en matière d'eGovernment. En effet, le 26 juin 2013, la directive européenne PSI de 2003 encourageant la diffusion des *open data*, a fait l'objet d'une mise à jour à transposer par les États-membres pour le 18 juillet 2015 (Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013) .

Comment se préparer le cas échéant à ces adaptations ? Quelles applications réalistes peut-on envisager et comment exploiter au mieux existant ? Quelles sont les questions ouvertes résiduelles ? En vue de proposer des pistes de réponses à ces interrogations, nous avons réalisé

une étude sur la question et une « *research note* » à ce sujet sera publiée sur ce site dans les semaines qui viennent. Un point crucial y occupe une place stratégique : la qualité des *open data* eu égard à leurs usages attendus, qualité potentiellement douteuse, dans la mesure où celles-ci prolifèrent dans l'environnement ouvert et non contrôlé du web.

En guise d'introduction à la prochaine mise en ligne de ce document de recherche, nous proposons ici de démystifier quelques idées reçues à propos des *open data*, et ce faisant, de dégager plusieurs bonnes pratiques pour en maintenir la qualité dans le temps.

**« Open » ne signifie pas nécessairement « gratuit »**



Certains mouvements en charge de la promotion de l'*open data* estiment que toutes les données ouvertes devraient être gratuites et militent dans ce sens. Sur le plan financier, la directive européenne PSI 2013 autorise toutefois la mise à disposition d'*open data* publiques moyennant un coût marginal optionnel (correspondant au coût de production et de diffusion des données). La directive laisse cependant ouvertes certaines interprétations, ne rendant pas toujours aisées les modalités de calcul.

Même si la question fait débat, le fait de diffuser des données payantes peut constituer une garantie de qualité. À titre d'exemple, en Belgique, la Banque Carrefour des Entreprises (KBO), source authentique des entreprises belges, prévoit, en plus de la diffusion gratuite d'un sous-ensemble d'informations, la diffusion payante de certaines des données dont elle a la gestion. Celles-ci seront notamment accessibles via un service web reposant sur les standards XML et SOAP. Un enregistrement obligatoire des utilisateurs est requis (ce qui permet éventuellement de récolter leurs commentaires sur la qualité des données ainsi que toutes leurs questions à ce propos). Le phénomène des *open data* et du

*crowdsourcing* peut ainsi être exploité en vue d'améliorer la qualité de sources mises à disposition du public dans le cadre d'une « gestion collaborative des anomalies » par plusieurs communautés d'utilisateurs potentiellement intéressés par les mêmes données.

On observe par ailleurs depuis peu une volonté de diffusion de données payantes dans le secteur privé bancaire, s'agissant par exemple des banques BNP Paribas Fortis et ING. Ces évolutions récentes suscitent, quant à elles, quelques inquiétudes en matière de *privacy*... en effet, « *open* » ne signifie pas nécessairement « libre de droit ».

### **« *Open* » ne signifie pas nécessairement « libre de droit »**

La directive européenne PSI de 2013 renforce l'obligation de transparence sur les conditions de licence. Selon les enjeux de l'information et de leurs usages, celles-ci pourront couvrir, notamment :

- la question du *copyright* et de la propriété intellectuelle;
- les risques liés au non-respect de la vie privée, même si les données ont été anonymisées (plusieurs statistiques croisées entre elles pouvant contribuer à l'identification d'individus, à travers un phénomène de « triangulation ») ;
- et enfin, les risques liés à l'usage inadéquat d'information de qualité incertaine. Ainsi le projet « *Citadel on the move* », qui encourage explicitement à publier des données ouvertes, quelle que soit la qualité de la source (« *accept mistakes : no dataset is perfect* ») se couvre avec une licence appropriée (« *Standard Open Licences such as CCBY and CC-O have watertight legal protection against liability for the accuracy of data. For more information see : <https://creativecommons.org/licences>* »).

Afin de déterminer la licence adéquate, on pourra en effet puiser dans les licences offertes par les *creative commons*, les licences les plus restrictives interdisant par exemple la transformation du contenu diffusé. Au niveau fédéral, le site de l'ASA/DAV mentionné plus haut propose une licence type pour la réutilisation des données du secteur public, dans la ligne de la nouvelle directive PSI 2013 et compatible avec toute licence libre qui exige a minima la mention de la paternité.



L'examen de cette question juridique des licences peut contribuer à améliorer la qualité de l'information. En effet, avant de penser à la production de nouvelles plateformes de « données ouvertes », il sera indispensable de recenser et de cartographier toutes les données publiques déjà mises à disposition à ce jour sur les sites d'eGovernment de chaque État-membre, d'en vérifier la qualité, l'usage et la conformité aux licences.

De nombreuses données ouvertes existantes ne sont en effet pas suffisamment connues ou ne se trouvent pas sous un format ou une licence adéquats. Certaines d'entre elles ne sont pas utilisées et sont laissées en friche, leur qualité se détériorant inévitablement avec le temps (ce qu'illustre l'adage « *use it or lose it* »). On pourra également découvrir à l'occasion de cet inventaire de nombreuses données « ouvertes » redondantes et mises à jour à des rythmes différents, dont l'exploitation demande a posteriori de nouvelles opérations de tests et de correction (mécanisme qu'illustre le concept bien connu de « *ghost factory* », désignant le temps et l'argent consacrés inutilement à produire des problèmes et à les corriger au sein d'une même entité). Un inventaire continu de cet ensemble permettra d'en améliorer et d'en maintenir la qualité dans le temps.

En effet, la multiplication des plateformes *open data* hétérogènes couvrant un même sujet avec la redondance d'information associée est à la source d'une dégradation de la qualité de l'information et de coûts supplémentaires inutiles. On observe ce phénomène non seulement au niveau local et national mais aussi au niveau européen où les plateformes d'*open data* sont multiples.

### **“Open” ne signifie pas nécessairement “public”**

Une fois la question des licences traitée, il est possible de combiner les *open data*, en vue de les diffuser soit publiquement, soit dans un cercle restreint. Alors que les « administrés » (entreprises, citoyens, employeurs, etc.) ne cessent de communiquer de l'information aux administrations, l'idée

est que ces dernières leur fournissent également en retour des informations à valeur ajoutée (en plus de leur mission première, en tant qu'élément constitutif de l'appareil d'État : prélèvement et redistribution des contributions, exécution de services au profit des administrés, application de la loi, etc.). Par ailleurs, à côté de la directive PSI 2013, dans l'intérêt commun opérationnel d'un domaine d'application donné, des informations peuvent être conçues de façon à être ouvertes dans le cadre d'un réseau secondaire « fermé ». On parlera alors « pragmatiquement » (et non juridiquement) de « *closed data* » interopérables entre elles.

Ce type d'application demandant un croisement entre sources préexistantes, il est utile de rappeler, même si cela relève du bon sens, les principes de gouvernance et de bonnes pratiques conceptuelles et fonctionnelles lors de la mise en place de toute nouvelle application. Outre l'avantage immédiat, les informations correspondantes seront d'autant plus interopérables dans le cadre de toute nouvelle application de type « *open government* », quels que soient ses usages ultérieurs.

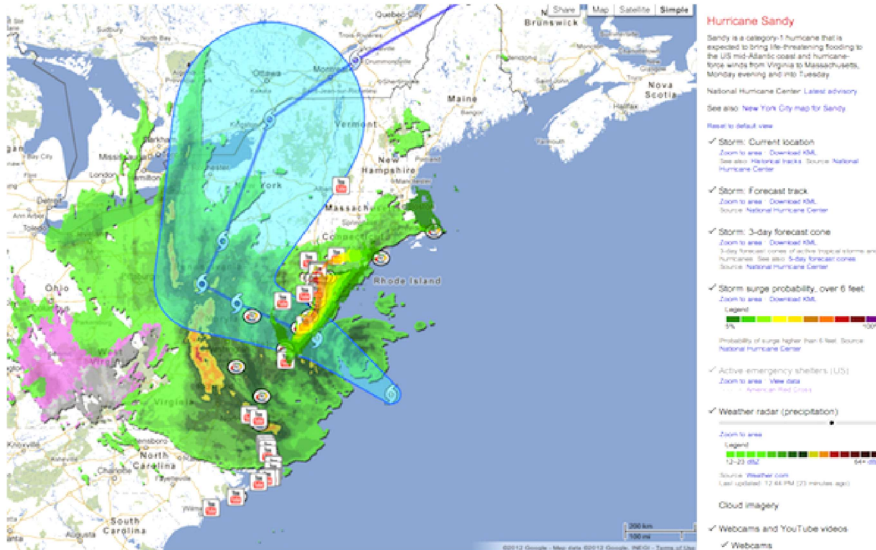
La problématique de la qualité de l'information prend en effet naissance en T0, lors de la phase de conception. Dans le cadre de la « *research note* » annoncée au seuil de ce post, plusieurs points importants seront rappelés à cet égard (standards, *data quality tools*, ...).

### **“Open” ne signifie pas nécessairement “sans contrôle”**

Progressivement associées dans le cadre de plateformes adéquates, les *open data* ont pour finalité, sur le plan applicatif de l'eGovernment, de donner le jour à des *mashups* et par exemple, à des services de « *citizen engagement* ».

Ainsi l'application de suivi de l'ouragan Sandy survenu aux USA en 2012, incluant des données publiques météorologiques mais aussi les constatations des citoyens et des propositions d'aide bénévole, constitue

un exemple remarquable dans ce sens.



L'application repose sur le croisement contrôlé de plusieurs temporalités, depuis les données continues de type « real time » relatives à la météo, aux données de gestion géographique ou statistiques, à la saisie d'informations purement empiriques reflétant le vécu des citoyens. De ce croisement émane une dynamique temporelle dont l'action sur le réel rétroagit à son tour sur le système d'information.

Tout au long de ce post, nous avons indiqué plusieurs pistes importantes susceptibles d'en maintenir la qualité dans le temps. Ce faisant, le mythe de la « *Tour de Babel* » ne se traduira pas nécessairement par l'écueil d'une confusion sans issue des langages et pourra peut-être donner le jour à une exploitation semi-contrôlée, riche et créative de la diversité des langues.



data quality

egov

information management

open data

---

MORE POSTS

## Je data beschermen tegen beheerders: 'on-premise' Confidential Computing

2026-03

## Protéger ses données des administrateurs : l'informatique confidentielle « on-premise »

2026-03

## De performance van LLM's: Een vergelijkende analyse tussen Frans en Nederlands

2026-03

## Made by Smals Research – Privacyvriendelijk Kruisen van Persoonsgegevens

2026-02

Search

Search

Newsletter & webinars:

Dutch  French

Your email address

Subscribe

**Keywords:**

analytics artificial intelligence big data blockchain bpm chatbot

cloud computing cost cutting cryptography data center

data quality development eda egov event gis

information management machine learning managing it costs

methodology mobile natural language processing open source privacy

productivity security social software design

software engineering standards

**Smals Research**

© Smals Research – License/Disclaimer: [FR](#) / [NL](#)