

MEILLEURES PRATIQUES RELATIVES AUX DONNÉES OUVERTES

Bart Hanssens (BOSA)

TASK FORCE OPEN DATA opendata@bosa.fgov.be

04/2023

Table des matières

Introduction.....	2
Les données ouvertes, de quoi s'agit-il ?	2
Directive européenne sur les données ouvertes et transposition en droit belge	2
La stratégie fédérale relative aux données ouvertes	3
La Task Force « Open Data »	3
Le portail fédéral des données ouvertes, data.gov.be.....	3
Directives pratiques.....	4
Adjudications et transfert de droits de propriété.....	4
Données ouvertes et gestion de données en général.....	4
Qualité des données.....	5
Interprétation des données.....	5
Conditions d'utilisation et licences.....	6
Calcul et imputation de redevances.....	7
Publication et mention sur data.gov.be	7
Formats des fichiers	8
Ouverture de données par le biais d'API/de services web	8
Coordonnées	9
Répondre aux questions des citoyens et des entreprises.....	9
(Ré)utiliser soi-même des données ouvertes.....	10
Informations supplémentaires.....	11
Liens.....	11

Introduction

Les données ouvertes, de quoi s'agit-il ?

Les données ouvertes sont des données qui :

- ne contiennent pas (ou plus) d'informations à caractère personnel ou d'autres informations sensibles,
- sont facilement réutilisables dans des processus automatiques,
- peuvent être réutilisées gratuitement (aussi à des fins commerciales).

Généralement, ces données sont publiées par des services publics mais elles peuvent aussi l'être par des entreprises, des institutions académiques ou des particuliers.

En fait, les données ouvertes ne sont pas nouvelles. Les autorités publient depuis des années des statistiques (par exemple) qui sont réutilisées par des journalistes, des chercheurs ou des entreprises. Cependant, depuis 2010 environ, elles ont été propulsées au devant de la scène grâce à une série d'initiatives des États-Unis et du Royaume-Uni notamment.

Directive européenne sur les données ouvertes et transposition en droit belge

Depuis un certain temps déjà, l'Europe disposait d'une directive sur l'accessibilité des données des services publics (informations du secteur public) afin de promouvoir la transparence des autorités publiques. Cette directive a été refondue afin d'encourager la réutilisation de ces données, au maximum pour une somme modique et de préférence gratuitement.

Le raisonnement est que les services publics travaillent avec l'argent des contribuables, de sorte que la collecte des données est déjà payée par les citoyens et les entreprises. Dans la mesure où ces données ne contiennent pas de données à caractère personnel ou d'autres données sensibles, il est donc logique que les contribuables y aient accès et puissent créer eux-mêmes des applications.

En effet, les entreprises qui réutilisent des données pour créer des applications performantes créent une valeur ajoutée économique et, en principe, paient à nouveau des impôts sur les bénéfices supplémentaires. D'autres parties, telles que des universitaires ou des citoyens engagés, peuvent réaliser des analyses ou des applications ayant une valeur ajoutée éducative ou sociale.

Au niveau fédéral, la deuxième version de la directive européenne ISP a été transposée dans la « Loi du 4 mai 2016 relative à la réutilisation des informations du secteur public » et dans « L'Arrêté royal du 2 juin 2019 relatif à la réutilisation des informations du secteur public ». La transposition de la troisième version est en préparation.

Pour plus d'informations, veuillez consulter

- <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>
- http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2016050417&table_name=loi
- http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2019060205&table_name=loi

La stratégie fédérale relative aux données ouvertes

Par ailleurs, les autorités fédérales ont défini en 2015 une stratégie relative aux données ouvertes, qui encourage à la fois les initiatives axées sur la demande que les initiatives proactives. Le seuil de réutilisation doit également être maintenu aussi bas que possible.

La Task Force « Open Data »

Afin de soutenir les services publics fédéraux, l'A.R. a mis en place une Task Force, composée de membres de la DG Transformation digitale du SPF BOSA et de l'Agence pour la Simplification Administrative.

Ils ont notamment aidé les services publics fédéraux en rendant des avis juridiques et techniques et ont collaboré avec des experts en données ouvertes des Régions et de la Commission européenne. En outre, la Task Force est en contact avec « OpenKnowledge Belgium », la communauté belge de données ouvertes au sein de laquelle sont représentés des universitaires et des entreprises.

Par ailleurs, ils répondent aux questions simples des citoyens et des entreprises et transmettent les questions plus détaillées au(x) service(s) public(s) le(s) mieux placé(s) pour y donner suite. Il peut s'agir tant de services publics fédéraux que régionaux ou locaux.

La Task Force « Open Data » est joignable via le formulaire de contact disponible sur le portail fédéral des données ouvertes et par e-mail (opendata@bosa.fgov.be).

Le portail fédéral des données ouvertes, data.gov.be

Le portail fédéral des données ouvertes comprend un aperçu des références aux ensembles de données ouvertes belges, indépendamment de la partie qui publie ces données (État fédéral, provinces, Régions, communes, intercommunales, entreprises...) ou de l'endroit où les fichiers ou services web sont gérés.

L'accent est mis ici sur les « références » ou les liens : les fichiers de données ouvertes ne sont donc pas copiés sur une plate-forme centrale unique. Dans la pratique, cela ne serait pas réalisable vu que les données proviennent de différents niveaux de pouvoir et que, dans certains cas, ces fichiers doivent (aussi) déjà être obligatoirement mis à disposition par le biais d'autres portails et plates-formes.

Cependant, la structure des métadonnées (titre, description, liens, thèmes...) est au maximum alignée selon la norme européenne DCAT-AP. En effet, ces métadonnées sont également partagées avec le portail européen des données ouvertes, qui répertorie tous les ensembles de données ouvertes de l'ensemble des États membres de l'U.E.

Les citoyens et les entreprises, mais aussi les services publics, peuvent chercher des ensembles de données sur le portail et poser des questions par le biais du formulaire de contact. Par ailleurs, le portail fédéral contient aussi de la documentation, des présentations et des rapports supplémentaires de groupes de travail.

- <https://data.gov.be/fr/info-faq>

Directives pratiques

Adjudications et transfert de droits de propriété

Les services publics ne peuvent évidemment pas publier de données sous forme de données ouvertes dont ils ne sont pas propriétaires. Les données achetées ou les données d'autres services publics ne peuvent donc pas être publiées comme bon leur semble.

Cette règle s'applique également aux données produites ou collectées dans le cadre d'un marché public : à l'issue du marché, le service public ne devient pas automatiquement propriétaire des données, y compris les données « brutes », qui sont souvent nécessaires pour parvenir au produit final ou au service. Sans disposition explicite, ces données sont présumées rester la propriété de l'adjudicataire.

Recommandation : veillez à ce que les marchés publics prévoient une clause stipulant explicitement que les droits sur les produits et services, et sur toutes les données « brutes » collectées ou produites pour réaliser les produits et services, sont transférés au service public adjudicateur.

Données ouvertes et gestion de données en général

Comme déjà indiqué, les données ouvertes ne sont pas vraiment nouvelles. De nombreux services publics sont déjà familiarisés avec la publication et la gestion de données en recourant à certaines normes (internationales).

Pour d'autres services, les « données ouvertes » peuvent mener à des questions et à des discussions qui n'ont en fait aucun lien avec l'aspect « ouvert » mais bien avec la gestion des données de manière générale. « Qui est le propriétaire des données ? Qui peut en dire plus sur le contenu et sur la façon dont ces données ont été collectées ? ». Il s'agit de questions que l'on peut, et même que l'on doit, se poser si l'on (ré)utilise des données de manière générale, même s'il s'agit de la réutilisation au sein du propre service public.

Recommandation : soyez critique mais constructif. L'ouverture des données peut être un premier pas vers un meilleur traitement des données en général.

Qualité des données

Parfois, la « qualité insuffisante » est invoquée comme raison pour ne pas mettre à disposition des données sous forme de données ouvertes.

Il est cependant difficile de justifier pourquoi les données ne sont pas assez bonnes pour être publiées, mais assez bonnes pour continuer à être utilisées en interne... La qualité peut également varier d'une application à l'autre : un plan des rues incomplet peut suffire pour un guide touristique, mais peut-être pas pour la livraison de colis. Il relève dès lors de la responsabilité de la partie qui réutilise les données d'en évaluer la qualité et de juger si elle est suffisante pour son application.

Recommandation : publiez les données et mentionnez (si possible) brièvement les limitations connues. Établissez par ailleurs un plan afin d'améliorer la qualité si nécessaire, et soyez ouvert à d'éventuelles questions et suggestions de parties externes.

Interprétation des données

Le fait que « des externes sont incapables d'interpréter correctement ces données » est souvent invoqué comme raison pour ne pas publier des données sous forme de données ouvertes.

Les autorités disposent en effet de beaucoup d'expertise en interne mais il est aussi nécessaire de documenter plus clairement cette expertise afin qu'elle puisse au moins être transmise aux nouveaux collaborateurs sans que la continuité des services ne soit menacée. Et, vu que cette documentation est de toute façon disponible, autant la partager avec le grand public... Cependant, en dehors des autorités également, nombreux sont les experts qui peuvent peut-être examiner les données sous un autre angle, ce qui peut mener à de nouvelles idées ou applications.

Recommandation : publiez les données et précisez (si nécessaire) comment les données ont été collectées et comment elles doivent être interprétées.

Conditions d'utilisation et licences

L'Arrêté royal prévoit une série de possibilités concernant les conditions d'utilisation par le biais d'un modèle « par étapes ».

En règle générale, aucune condition n'est imposée : la réutilisation gratuite est donc autorisée, également à des fins commerciales, et il n'est pas obligatoire de mentionner la source. On peut comparer cela à la déclaration connue internationalement, « Creative Commons Zero (CC0) ».

Il est possible d'y déroger si une bonne raison est mentionnée sur le site web du service public et si cette motivation a été envoyée au préalable à la Task Force « Open Data » à des fins d'avis (dans le cas contraire, la licence n'est pas valable).

Une première dérogation prévue est la possibilité d'obliger de mentionner la source, ce qui équivaut à la licence internationale 'Creative Commons Attribution (CC-BY) ». Conformément à cette licence, la réutilisation est donc toujours gratuite, même pour les applications commerciales.

Si cela ne suffit pas, un service public peut exceptionnellement choisir d'imputer des « coûts marginaux ». L'A.R. prévoit une énumération du type de coûts qu'il est permis d'imputer mais ne comprend pas une liste des montants exacts. Par conséquent, s'il n'existe aucune autre loi ou aucun autre arrêté fixant un tarif fixe pour un ensemble de données défini, le service public doit lui-même faire le calcul pour déterminer le montant exact et ensuite publier ce montant.

Cette option est généralement déconseillée : dans de nombreux cas, les frais administratifs en la matière (calcul, envoi d'une proposition, facturation...) sont supérieurs aux bénéfices potentiels de la vente des données.

Si cela non plus ne suffit pas, il existe encore la possibilité de rédiger une autre licence, mais à titre exceptionnel. Cette méthode est aussi généralement déconseillée : des licences très détaillées ont parfois été rédigées dans le passé, et soit elles ne différaient pas beaucoup des licences Creative Commons connues internationalement, soit elles n'avaient pas été contrôlées ou imposées dans la pratique.

La rédaction correcte d'une propre licence demande du temps et des efforts à un service public et cette licence n'est pas non plus conviviale pour les parties qui réutilisent les données, qui veulent souvent combiner des données de différentes sources.

Recommandation : utilisez autant que possible la déclaration « Creative Commons Zero (CC0) » ou (après motivation et moyennant l'avis positif de la Task Force) la « Creative Commons Attribution (CC-BY) ». Évitez des licences propres.

Calcul et imputation de redevances

Il n'est pas obligatoire, mais bien autorisé, d'imputer une « redevance marginale » pour la fourniture de données, et, dans quelques cas spécifiques, le coût a été fixé dans un Arrêté royal ou dans un Arrêté ministériel.

Dans les autres cas, un calcul détaillé doit pouvoir être soumis afin que les frais liés au calcul et à la facturation puissent finalement être supérieurs aux bénéfices potentiels. En outre, la majorité des demandes proviennent de citoyens et d'entreprises qui paient des impôts et qui fournissent donc déjà une contribution financière au fonctionnement des autorités. Les données doivent dès lors être publiées autant que possible gratuitement.

Ce raisonnement vaut aussi dans le sens inverse : ce n'est pas parce qu'une entreprise peut gagner de l'argent avec des services ou combinaisons de données basées sur des données ouvertes que les autorités sont autorisées à imputer des coûts à cette fin.

Sauf dans des cas très spécifiques, la conclusion de contrats d'exclusivité est interdite.

Recommandation : n'imputez pas de coûts pour la mise à disposition de données. Si cela est néanmoins absolument nécessaire, publiez le calcul détaillé afin de justifier le prix.

Publication et mention sur data.gov.be

Le portail fédéral des données ouvertes, data.gov.be, rassemble des liens et d'autres métadonnées de fichiers de données ouvertes. Les fichiers eux-mêmes ne sont pas copiés mais restent sur le site web ou le portail des services publics.

L'objectif n'est évidemment pas de réintroduire manuellement sur data.gov.be les liens, titres, etc., qui ont déjà été décrits. Ces informations doivent en effet pouvoir être reprises le plus possible de manière automatique.

Pour ce faire, la méthode privilégiée consiste à créer et à échanger un fichier de métadonnées « DCAT-AP », mais, selon le système dans lequel les données sont gérées ou selon le système avec lequel le site web a été conçu, il sera également possible de recourir à une API ou à un flux d'actualités (*newsfeed*) et, dans le pire des cas, à extraire (« moissonner ») une page d'aperçu sur laquelle tous les ensembles de données ouvertes sont répertoriés. Contactez la Task Force « Open Data » afin qu'elle examine avec vos collaborateurs techniques comment réaliser cette connexion.

Recommandation : contactez la Task Force « Open Data » pour convenir de quelle manière les liens vers les ensembles de données doivent être (automatiquement) publiés et mis à jour sur le portail fédéral des données ouvertes.

Formats des fichiers

Il n'existe pas de « format générique de données ouvertes », tout comme il n'en existe pas pour toutes les applications.

Des statistiques sont par exemple publiées au format XLSX ou CSV. Pour des informations géographiques, on utilise souvent ShapeFile, et GTFS est populaire pour les horaires des transports en commun.

Il est important que les données soient disponibles dans un format que l'on peut facilement traiter de manière automatique : il est en effet très difficile pour des machines de récupérer des données de manière structurée d'un fichier PDF ou DOCX comprenant des images et des tableaux ou de cartes comportant uniquement des codes couleurs.

Il est naturellement toujours permis de publier ces rapports, cartes ou infographiques mais veillez alors à ce que les données « brutes » soient disponibles dans un format tel que CSV, JSON, KML, ShapeFile, sqlite, TTL, XLSX ou XML.

Recommandation : publiez les données « brutes » sur lesquels les rapports et les cartes sont basés également dans des formats connus qui peuvent être traités facilement de manière automatique.

Ouverture de données par le biais d'API/de services web

Les utilisateurs se renseignent non seulement sur les fichiers téléchargeables mais aussi parfois sur la disponibilité d'API ou de services web afin que les données puissent être intégrées plus facilement dans des applications ou dans leurs propres systèmes.

C'est surtout important pour les données qui changent très souvent, ou pour de grands ensembles de données dont l'utilisateur ne veut utiliser qu'une petite partie, comme les retards de train ou des données géographiques. Comme il n'est pas réalisable – ni nécessaire – de fournir une API pour tous les ensembles de données possibles, il faudra toujours trouver un compromis entre le coût de développement et de maintenance de l'API et la valeur ajoutée ou les économies que les utilisateurs peuvent réaliser en l'utilisant.

De manière générale, il est utile d'offrir une API si un site web propose déjà un écran de recherche permettant aux utilisateurs d'effectuer manuellement des recherches. Dans certains cas, on peut faire tourner l'API et l'écran de recherche par le biais de canaux distincts afin que l'écran de recherche continue à fonctionner si l'API devait être (trop) lourdement chargée.

Dans de nombreux cas, on recommande aussi toujours d'offrir régulièrement, outre l'API, un fichier complet à télécharger afin que l'utilisateur puisse choisir. Un bon exemple : les statistiques du site infocenter.belgium.be, qui sont disponibles tant sous forme téléchargeable que par le biais d'une API.

Recommandation : quand vous développez de nouveaux systèmes, envisagez la possibilité de prévoir une API afin que les utilisateurs puissent l'intégrer dans leurs propres systèmes. En règle générale : si une banque de données en ligne dispose d'un écran de recherche pour rechercher manuellement des données, il est utile de prévoir aussi une API afin de pouvoir automatiser les recherches.

Coordonnées

Comme pour les brochures classiques ou d'autres informations sur les sites web des autorités, il faut savoir clairement pour chaque ensemble de données par le biais de quel canal les utilisateurs peuvent poser des questions.

Il est dès lors préférable de publier une adresse e-mail à l'endroit où les ensembles de données sont mis à disposition. Utilisez à cette fin une adresse générale ou un groupe de messagerie au lieu d'une seule personne spécifique afin que les questions éventuelles puissent également faire l'objet d'un suivi pendant les vacances ou les absences. Il peut s'agir de la même adresse que celle utilisée pour d'autres questions.

Une alternative est de prévoir un lien vers un formulaire de contact général ou spécifique.

Recommandation : veillez à disposer d'une adresse e-mail à laquelle les citoyens, les entreprises et d'autres services publics peuvent envoyer leurs question sur les ensembles de données.

Répondre aux questions des citoyens et des entreprises

Les citoyens et les entreprises (ou autres services publics) peuvent donc poser des questions sur des données déjà publiées, ou peuvent rechercher des données qu'ils présument être en possession de votre service.

Il est très probable que votre service ne dispose pas de ces données : faites-le savoir au plus vite au demandeur. Si la demande n'est en fait pas destinée à votre service, essayez de la transmettre vous-même au service approprié (il peut aussi s'agir d'un service d'une autorité régionale ou locale) ou transférez la question à la Task Force fédérale « Open Data » (opendata@bosa.fgov.be) et informez le demandeur de l'instance à laquelle vous avez transmis sa requête.

Informez aussi le demandeur si sa requête n'est pas claire d'emblée ou si le délai de réponse peut être long. Selon l'A.R., une réponse doit en effet être communiquée dans un délai de 10 jours ouvrables (30 en cas de demande complexe), mais un bref accusé de réception donne une image professionnelle.

La réponse peut bien entendu également être négative : si les données ne sont pas la propriété des autorités, contiennent des données personnelles ou ne peuvent être libérées pour des raisons de sécurité, le demandeur ne peut pas les obtenir. Dans ce cas, il doit cependant être averti de la raison pour laquelle les données ne sont pas libérées.

Les autorités ne sont pas obligées de faire des efforts « déraisonnables » : le but n'est pas de procéder à des traitements très spécifiques pour arriver à un ensemble de données personnalisé.

Recommandation : envoyez aux entreprises et aux citoyens un bref accusé de réception. Traitez la demande (si possible) dans un délai de 10 jours ouvrables et mentionnez la raison en cas de réponse négative. Si la demande n'est pas destinée à votre service, transférez-la au service adéquat ou à opendata@bosa.fgov.be.

(Ré)utiliser soi-même des données ouvertes

Les services publics ne publient pas uniquement des données ouvertes. Ils peuvent à leur tour également réutiliser des données ouvertes provenant d'autres services.

En voici un bel exemple : les fichiers d'adresses « BeST » (Belgian Streets). En collaboration avec les 3 Régions, les autorités fédérales publient une série de fichiers comprenant tous les noms de rues et les numéros de maison valables sous forme de données ouvertes, et ces fichiers sont à leur tour utilisés pour offrir des services web aux services publics fédéraux.

Recommandation : la réutilisation de données ouvertes par les services publics est fortement encouragée. Bien que ce ne soit pas obligatoire, il est recommandé de contacter le service public qui publie ces données. Il aura ainsi également une meilleure vue de l'utilisation et des éventuels besoins spécifiques.

Informations supplémentaires

Liens

Belgique

- Task Force « Open Data » : opendata@bosa.fgov.be
- Portail fédéral des données ouvertes : <https://data.gov.be>
- Rapports et présentations Groupe de travail Portails : <https://data.gov.be/fr/dcat-groupe-de-travail>
- Loi relative à la réutilisation des informations du secteur public : http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2016050417&table_name=loi
- Arrêté royal relatif à la réutilisation des informations du secteur public : http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=2019060205&table_name=loi

Europe :

- Portail européen de données : <https://www.europeandataportal.eu/fr>
- Directive européenne relative aux données ouvertes : <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>