

MATTHIEU BRIENT,
chef de projet à OpenDataFrance



L'OPEN DATA CONSTITUE UNE RÉPONSE À LA MAÎTRISE DE L'IMPACT ENVIRONNEMENTAL DU NUMÉRIQUE

Arabesques : quelle est l'origine du référentiel Green Data d'OpenDataFrance¹ ?

MATTHIEU BRIENT : ce travail a été impulsé par le conseil d'administration d'OpenDataFrance, composé d'élus de collectivités territoriales, qui a choisi de consacrer sa feuille de route notamment à la responsabilité des données et à la sobriété numérique. Des travaux importants avaient déjà été menés par la mission interministérielle numérique écoresponsable et par l'Institut du numérique responsable mais concernant la data, les recommandations restaient très générales. Le référentiel Green Data est un guide de bonnes pratiques qui permet de mieux comprendre l'impact des données ouvertes et des services qui y sont associés pour pouvoir les limiter.

Arabesques : dans quel contexte législatif s'inscrit-il ?

M.B. : il existe un cadre juridique qui impose aux institutions et collectivités de prendre en compte l'impact environnemental du numérique, notamment par la loi REEN de novembre 2021 visant à réduire l'impact environnemental du numérique, tandis que la loi Pour une république numérique d'octobre 2016 impose aux opérateurs publics l'ouverture des données qu'ils produisent. Sachant que seuls 15 % des collectivités ont ouvert pour l'instant leurs données, on peut s'attendre à un accroissement exponentiel de la volumétrie des données produites. Cette accumulation de données se répercute sur les besoins d'équipements et d'énergie. Ces besoins, malgré les gains d'efficacité notamment de la consommation énergétique des infrastructures, peinent à compenser pleinement l'accroissement des usages.

Arabesques : quelles sont les principales préconisations à retenir ?

M.B. : la première est d'établir des priorités dans le choix des données à publier. Aujourd'hui, le cadre légal impose l'ouverture de toute donnée dont disposent les collectivités publiques mais il existe aussi un socle commun des données locales qui fournit des standards aidant les collectivités à prioriser l'ouverture des données selon leur intérêt. Un volet de recommandations concerne la standardisation, la qualité et l'interopérabilité des données, indispensables à leur réutilisation. Une autre préconisation porte sur la volumétrie et la granularité des données, géographiques ou temporelles, afin qu'une collectivité souhaitant accéder à un jeu de données la concernant ne soit pas obligée de télécharger l'ensemble du corpus à l'échelle nationale. Une autre thématique que nous avons traitée est celle de l'accès à ces données. Actuellement elles sont publiées sur le portail open data de la collectivité, sur data.gouv.fr ou sur des portails thématiques, tourisme, transports. On s'est rendu compte que dans certains cas, les données étaient dupliquées et non moissonnées ce qui pose des questions concernant leur mise à jour. On incite donc les producteurs et utilisateurs de données à utiliser un accès *via API*, même si cet accès ne peut pas être exclusif pour des questions juridiques, car il permet le maintien et la mise à jour des données quel que soit leur lieu de publication et d'hébergement.

Arabesques : qu'en est-il de l'archivage de ces données ?

M.B. : c'est une question sur laquelle nous avons aussi beaucoup travaillé. Certaines données sont dépubliées car elles ne sont plus prioritaires mais doivent rester

accessibles pour une utilisation ultérieure. L'objectif serait de pouvoir plus facilement identifier les données obsolètes des données vivantes, et de travailler sur l'échantillonnage, c'est-à-dire ne conserver dans certains cas qu'un seul échantillon significatif, déterminé selon des modèles mathématiques, plutôt l'ensemble d'un jeu de données, afin de limiter la charge de stockage.

Arabesques : quels seront vos prochains axes de travail ?

M.B. : les collectivités territoriales produisent bien d'autres datas que les données ouvertes, et beaucoup plus volumineuses, comme les données en temps réel dans les *smart cities* ou les données géographiques. Nos recommandations sur l'open data pourraient être ajustées à ces autres types de données. Nous travaillons aussi sur la manière dont les données environnementales peuvent aider à la transition des territoires. Le recours de plus en plus fréquent à l'intelligence artificielle pour le traitement des données pose également question car ces modèles de *deep learning* vont mobiliser énormément de ressources, notamment énergétiques.

OpenDataFrance accompagne les territoires dans l'ouverture des données mais encourage fortement la rationalisation de la gestion de ces données, leur hébergement mutualisé, leur réutilisation pour éviter les duplications. Il est très important de noter que par l'ouverture et la mise à disposition des données *via* des portails bien référencés et mis à jour, l'open data constitue une réponse à la maîtrise de l'impact environnemental du numérique. ■

[1] <https://opendatafrance.gitbook.io/greendata-pour-un-impact-maitrise-des-donnees/greendata/preface-et-remerciements>