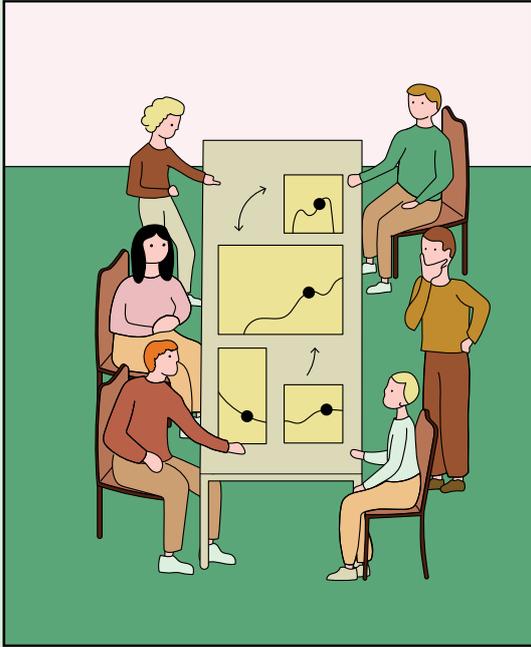
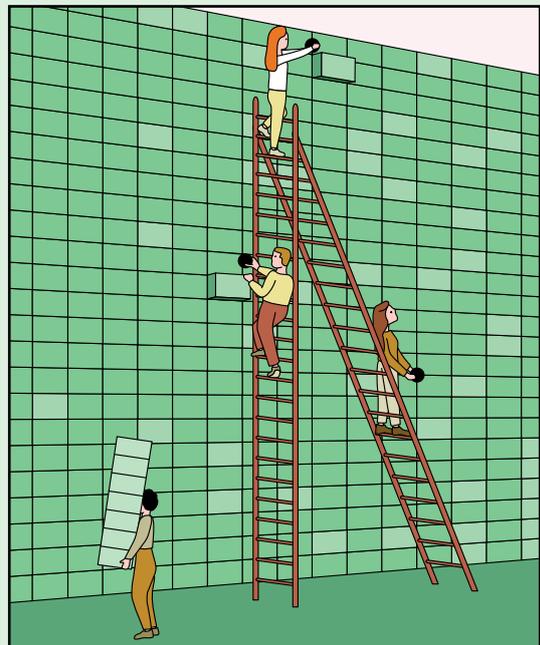
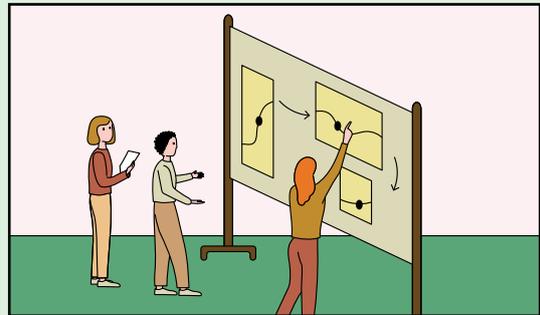
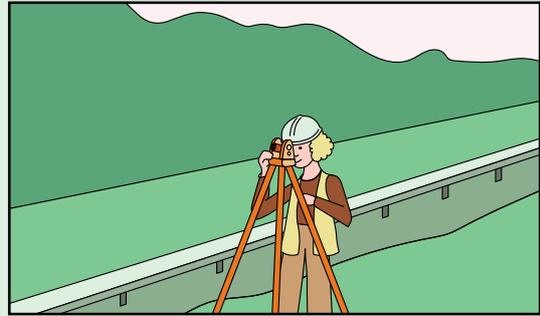


Référentiel GreenData



Pour un impact
environnemental
maîtrisé des données
ouvertes



Édition v1 - septembre 2022

Droits de reproduction, type de licence cc-by-sa

Rédaction : OpenDataFrance, Jean-Marie Bourgogne, Matthieu Brient
Création graphique et mise en page : Studio Bianccha, Natasha Stocchi
Illustrations : Natasha Stocchi

Référentiel GreenData

Pour un impact environnemental
maîtrisé des données ouvertes

REMERCIEMENTS

Depuis 10 ans, OpenDataFrance accompagne et fédère les collectivités et organisations investies dans l'ouverture des données publiques. Sa récente feuille de route stratégique vise à travers l'un de ses axes à s'intéresser aux enjeux de la responsabilité et de la sobriété numérique.

Une exigence appliquée à nous même

Comment mieux comprendre et limiter l'impact environnemental des données ? Comment contrôler les externalités négatives de leur ouverture, notamment lorsque leurs volumes ou leurs coûts de mise à disposition sont importants ? Comment penser et intégrer la soutenabilité des données et de leurs services associés ?

Partout, les volumes de données produites et hébergées sont déjà significatifs, et les projections anticipent une augmentation non négligeable. Conscient du rapport d'échelle entre certaines données ouvertes publiées et d'autres données plus volumineuses (vidéo, temps réel, géographiques...), ce travail est une première brique pour penser la sobriété des données à travers leur cycle de vie.

Ce guide s'articule aux nombreuses ressources produites par l'Institut du Numérique Responsable et la Mission interministérielle Numérique écoresponsable (MiNumEco). Il vise à enrichir et renforcer la sobriété des politiques numériques sur le volet data. Nous remercions ici les membres d'OpenDataFrance, représentant.e.s et agents des collectivités qui ont contribué à l'élaboration de ce premier guide de bonnes pratiques. Nous remercions aussi les partenaires et experts associés à cette production pour leurs éclairages.

En rationalisant la gestion des données, en encourageant l'ouverture et la réutilisation limitant ainsi les duplications, l'Open Data est aussi une réponse à la maîtrise de l'impact environnemental du numérique. A nous de nous saisir de cette opportunité.

Louise Vialard
Vice-présidente d'OpenDataFrance

Sommaire

—

INTRODUCTION	7
LES BONNES PRATIQUES	10
PLANIFICATION	12
BP1 - Engager et conduire une démarche “donnée ouverte et responsable”	14
BP2 - Prioriser les choix de données publiées	16
PRODUCTION	18
BP3 - Standardiser les jeux de données	20
BP4 - Documenter précisément les métadonnées	22
BP5 - Proposer une granularité temporelle et géographique adaptée	24
ANALYSE	26
PUBLICATION	28
BP6 - Réduire le volume unitaire des jeux de données	30
BP7 - Proposer un accès aux données par API	32
BP8 - Faciliter la découvrabilité des jeux de données	34
CONSERVATION	36
BP9 - Mettre en place une politique d’archivage	38
EXPLOITATION	40
BP10 - Contrôler l’hébergement des données	42
POSTFACE	44

CONTEXTE

Pourquoi et à qui s'adresse ce guide ?

Ce référentiel est destiné à accompagner les organisations (collectivités territoriales, institutions, associations...) dans leur maîtrise de l'impact environnemental de la donnée. Il offre des pistes dans l'élaboration de plans d'actions pour un numérique responsable, sur son volet data. Il est conçu et proposé par OpenDataFrance et ses partenaires.

L'objectif de ce guide est, d'une part, de sensibiliser les différents acteurs aux impacts de certaines pratiques rencontrées au sein des organisations et, d'autre part, de partager des exemples de bonnes pratiques numérique responsable qui peuvent participer à la réduction de l'empreinte environnementale du numérique à travers ce prisme de la donnée ouverte et de ses services associés.

Au sein des organisations visées, le numérique ne concerne pas uniquement la direction du numérique. D'autres services transverses comme les achats, les ressources humaines, la communication, les services généraux, etc. sont également impliqués.

Ce guide s'adresse à un large public en relation avec les systèmes d'information tels administrateurs systèmes et réseaux, développeurs, acheteurs, mais aussi naturellement leurs usagers, les directions dites « métiers » et en premier lieu les décideurs.

Une production collective

Pour réfléchir à ces enjeux, identifier les bonnes pratiques actuelles et futures pour une gestion responsable des données, nous avons associé les collectivités territoriales dans un travail approfondi de collecte, de documentation et de propositions. Collectivités, experts et membres d'OpenDataFrance, ont été associés à ces travaux à travers différentes étapes :

- une phase de collecte des besoins des collectivités, à travers un support d'enquêtes et une série d'entretiens.
- une phase d'analyse critique des externalités négatives de la publication, du stockage, et de l'utilisation de données. Des ressources scientifiques et des entretiens auprès d'experts ont pu nourrir cette phase.
- une phase d'élaboration de propositions de bonnes pratiques discutées à travers une série d'ateliers.
- une phase de mise en consultation et d'enrichissement des bonnes pratiques.

Une production à itérer

Ce sujet de la sobriété et de la maîtrise de l'empreinte environnementale des données, ainsi que les outils disponibles à la mesure des impacts sont encore émergents. Afin de s'adapter au mieux à des effets non anticipés mais aussi à l'évolution des connaissances et ressources disponibles sur ce sujet, il est indispensable de paramétrer un suivi et des évolutions en fonction du

contexte. Ce travail est une première brique pour penser la sobriété des données. Des itérations sont à proposer et à construire collectivement.

Pour des données ouvertes et maîtrisées

Quel numérique voulons-nous ?

Que ce soit pour les habitants usagers, agents des services ou les organisations elles-mêmes, le numérique déployé et utilisé aujourd'hui n'est pas soutenable. Les impacts éthiques, sociaux, environnementaux, ou encore de gouvernance, désormais largement connus, démontrent que ce numérique n'est pas adapté au monde de demain.

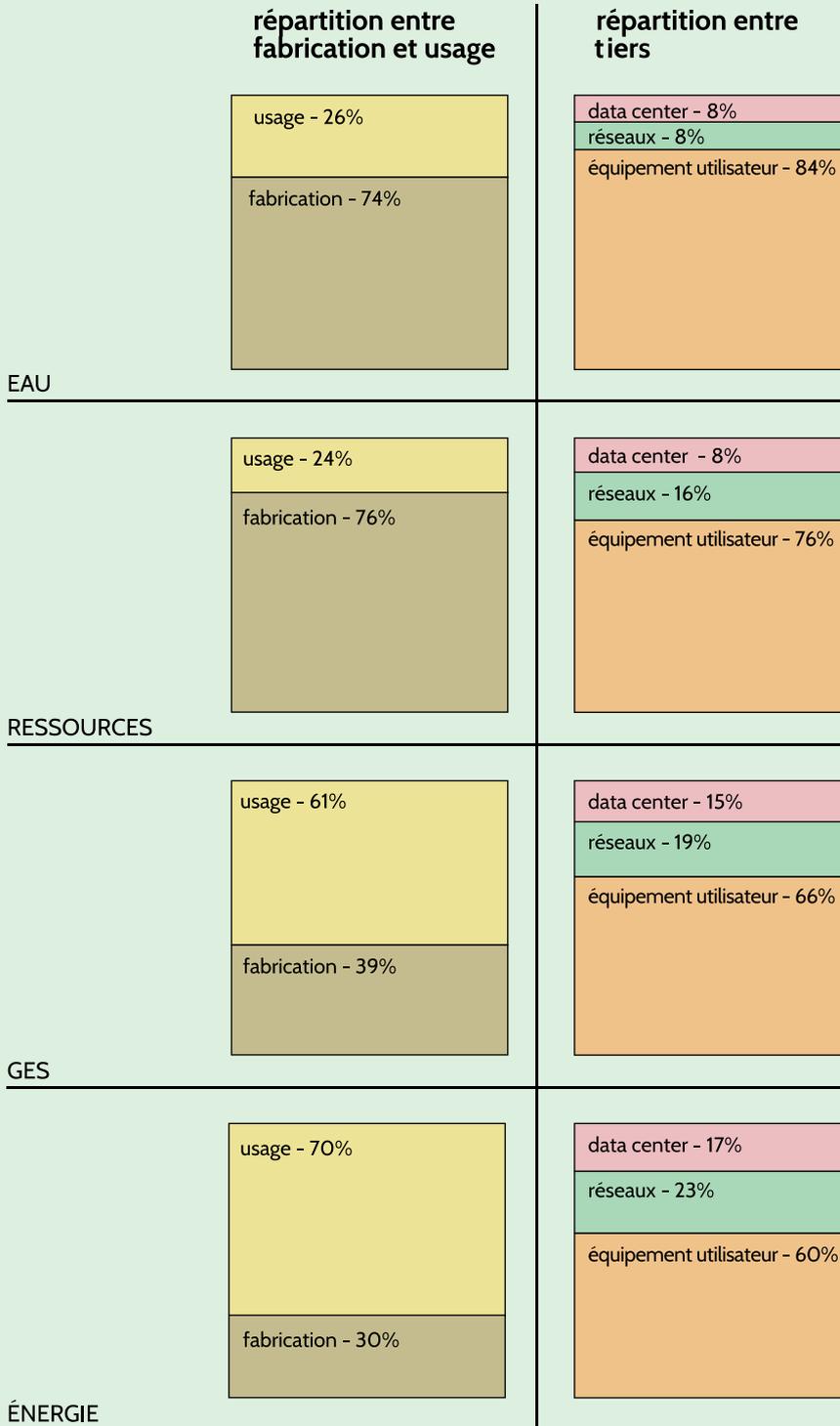
La transformation numérique de l'action publique est depuis longtemps à la recherche d'équilibres entre les gains de productivité et la qualité de services publics. Pour répondre à ces enjeux, il faut faire le choix d'innover pour longtemps et d'innover pour tous, en intégrant une forme de rationalité.

Nous le savons, le numérique (dont l'ouverture des données, sa mise à disposition et ses services associés), sont sources d'une empreinte environnementale non négligeable. Que ce soit majoritairement par les infrastructures et équipements nécessaires au déploiement de nouveaux services, ou encore aux ressources, dont l'eau et l'énergie, nécessaires à la disponibilité des services en continue, il est fondamental aujourd'hui d'aborder la question des services numériques en ayant une vision systémique de la maîtrise.

Ce travail propose d'explorer cette notion de maîtrise à trois niveaux: infrastructures et gestion de l'hébergement, la donnée et son cycle de vie, applications et services numériques proposés.

Ces trois dimensions identifiées offrent un large champ d'actions possibles. Nous proposons ici de les développer en 10 bonnes pratiques afin de maîtriser au mieux leurs impacts.

Répartition des impacts du numérique



IMPACT ENVIRONNEMENTAL DU NUMÉRIQUE,

ÉTAT DES LIEUX

Malgré l'opacité des méthodes et des données utilisées pour estimer l'impact environnemental du numérique, la littérature sur ce sujet s'accorde à affirmer que la matérialité du numérique impacte l'environnement sur de nombreux secteurs, et ce de manière croissante. En voici quelques-uns :

1. L'énergie: En 2021, le Shift Project évalue la part du numérique à 12% de la consommation totale d'électricité. La consommation des réseaux devrait selon NegaWatt augmenter de 10 % par an sur les 15 prochaines années. A noter que la consommation d'électricité nécessaire pour faire fonctionner tous les appareils et infrastructures numériques devrait augmenter jusqu'à 80 % entre 2020 et 2030 ([source](#)).

2. L'eau: 626 milliards de litres, c'est la quantité d'eau que les datacenters consommaient aux Etats-Unis en 2014. En France, peu de datacenters mobiliseraient un refroidissement par circuit d'eau. Cependant, l'empreinte eau de la production d'électricité française serait de 4 L / kWh. Un impact indirect mais bien réel des besoins en ressource des centres de données.

3. Les gaz à effet de serre: Le numérique représenterait 3 à 4 % des émissions de gaz à effet de serre (GES) dans le monde et 2,5 % de l'empreinte carbone nationale. Si rien n'est fait pour en réduire l'empreinte, ces émissions pourrait augmenter de manière significative: + 60 % d'ici à 2040, soit 6,7 % des émissions nationales ([pré-rapport de la mission d'information sur l'empreinte environnementale du numérique du Sénat](#)),

4. Les déchets: 53 mégatonnes (Mt) de Déchets d'Équipements Électriques et Électroniques (DEEE) ont été générés dans le monde en 2020, 12 Mt était attribuable à la zone Europe, à hauteur de 16,2 kg par personne (Vanessa Forti et al. «[The Global E-waste Monitor 2020: Quantities, flows and the circular economy potential](#)» (UNU). La production mondiale de déchets électroniques devrait augmenter de 100 % par rapport à 2014 et s'élever à 74 mégatonnes d'ici à 2030 ([source](#)).

Évolution du cadre légal en France

Le cadre légal français incite les acteurs publics et innovateurs à évaluer et maîtriser l'impact environnemental de leurs pratiques numériques.

Que ce soit à travers la Loi relative à la lutte contre le gaspillage et à l'économie circulaire (AGEC) ou la Loi visant à réduire l'empreinte environnementale du numérique en France (REEN), il convient de s'outiller pour mesurer et ainsi maîtriser l'empreinte environnementale des données, de ses infrastructures et services associés.

Au-delà de la sensibilisation, de la maîtrise du renouvellement d'appareils, 2 dispositions visées par la Loi REEN nous semble important à rappeler ici:

Favoriser les stratégies numériques responsables dans les collectivités – Les communes et les EPCI de plus de 50 000 habitants devront avoir d'ici 2024 défini une stratégie numérique indiquant les objectifs liés à la réduction de l'empreinte du numérique sur l'environnement, et les mesures mises en place pour les atteindre.

Promouvoir la sobriété des centres de données – Aller vers des centres de données et des réseaux moins énergivores. Faire émerger une régulation environnementale pour prévenir l'augmentation des consommations et émissions des réseaux et des centres de données. Cette disposition va dans le sens des Plans climat-air-énergie (PCAET) en intégrant par exemple le potentiel de récupération de chaleur des centres de données dans leur stratégie.

Ordre de grandeur de l'open data

La production de donnée au niveau mondial est exponentielle. Portée par une explosion des usages, en 2020, les données générées, copiées et consommées en ligne est estimée à plus de 60 Zettabytes (1 zetta représente environ 1 milliard de terabytes). Bien que ce volume soit essentiellement porté par une croissante forte du contenu vidéo et temps réel, la volumétrie seule n'est pas un indicateur suffisant d'impact du numérique.

En 2021, 14% des collectivités visées par la Loi République Numérique ont ouvert leurs données (voir [Résultats de l'Observatoire OpenData](#)). Environ 300 portails sont à ce jour ouverts et maintenus. Ils représenteraient un volume d'environ 30 To ([source](#) : [opendatarchives.fr](#)).

Conscient du rapport d'échelle entre certaines données ouvertes publiées et d'autres données plus volumineuses traitées par les collectivités (temps réel, géographiques, données d'opérateurs de services urbains...), se pose aussi aujourd'hui la question des traitements faits de ces données, notamment pour le développement d'IA. Depuis 2012, le nombre d'opérations nécessaire pour entraîner les modèles état de l'art de deep learning double tous les 3-4 mois ([source](#)). L'efficacité énergétique des data centers double tous les 3 ans seulement. D'où l'importance de réduire la taille des modèles et la taille des jeux de données.

La question ici n'est pas de savoir quoi ouvrir, mais comment mieux le faire. Pourquoi multiplier les fichiers liés à un même jeu de données, laisser des informations redondantes, répliquer des données publiées par ailleurs...

LES BONNES PRATIQUES

Avant tout, questionner le besoin

L'ouverture des données s'accélère. Elle répond ainsi aux obligations légales comme aux volontés politiques nationales et territoriales à des fins de transparence notamment.

Pourquoi s'intéresser à la sobriété des données et des usages ?

Une piste pour réduire significativement l'impact du numérique s'oriente dans ce référentiel à travers la notion de sobriété: **faire autrement, faire moins, voire ne pas faire.**

L'énergie requise à maintenir une infrastructure matérielle et réseau déjà déployée n'est que peu impactée par les usages de ces équipements (à moins d'éteindre ou rationaliser ces infrastructures).

Cependant, les terminaux et services hébergés dans des centres de données sont quant à eux corrélés aux usages numériques. De fait, la sobriété vise dans ce contexte à une forme de soutenabilité, à limiter le remplacement et l'empilement de technologies, ou encore l'effet rebond

Quelques éléments de réflexions à garder en tête tout au long du projet d'ouverture de données

Questionner le besoin pour évaluer la pertinence à publier certaines données.

En amont de l'ouverture des données, plusieurs questions doivent être posées afin d'évaluer collectivement la pertinence à produire et publier certaines données ou encore à développer de nouvelles fonctionnalités.

“Est-ce que cette donnée (ou ce service) est utile ? A qui ? A quelle fréquence ? Pourquoi ?”

Ces réflexions permettent de mieux cerner l'opportunité à développer de nouveaux services open data, à mieux comprendre les données essentielles qui les composent, leur niveau de détail (ou granularité géographique et temporelle), les modes de partage et d'actualisation (fichiers, API, restitutions).

L'ouverture « par défaut » (Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique) de toutes « les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental » reçues ou produites par la collectivité, peut s'accompagner de la mise en oeuvre de recommandations qui visent à privilégier, à prioriser, les données qui sont expressément demandées, en particulier lorsque le volume ou le coût de mise à disposition est important.

S'il y a demande, il y aura usage. Et si ces données respectent par ailleurs un standard, leur usage sera d'autant plus efficient.

Une approche itérative, pour une ouverture des données tirées par les usages

Plutôt que de s'appuyer sur un dispositif dont le coût écologique sera plus élevé que sa valeur d'usage, le développement de services open data et l'ouverture de données peut se faire de manière itérative. Cela permet de s'adapter en permanence, et au mieux, aux usages et au contrôle de ses impacts.

A titre d'exemple, la ville de Pirae, Polynésie française, a fait le choix de ne publier que les données standardisées.

Viser et entretenir une démarche d'amélioration continue

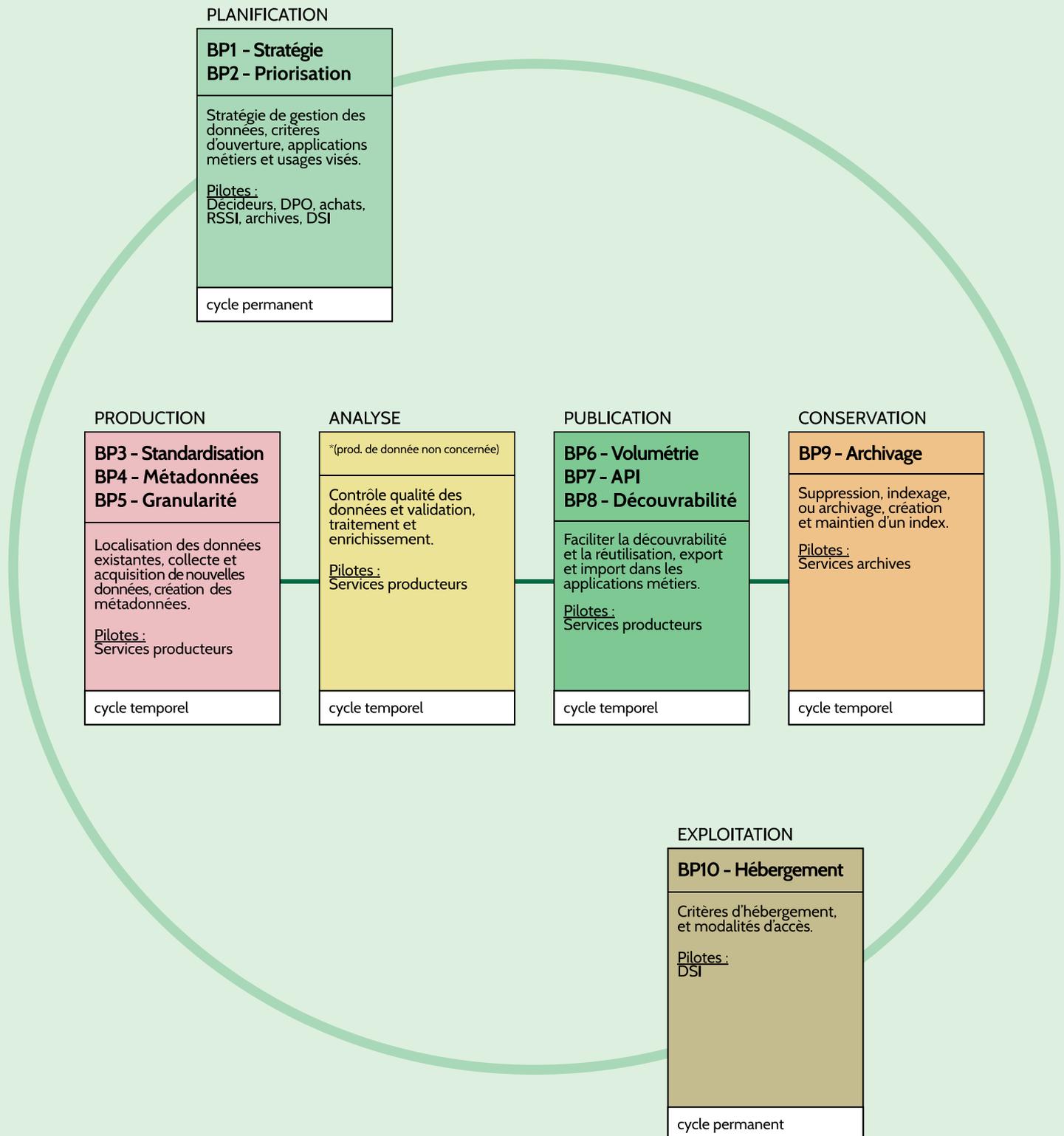
Ce sujet de la sobriété et de la maîtrise de l'empreinte environnementale du numérique, ainsi que les outils disponibles à la mesure des impacts sont encore émergents. Afin de s'adapter au mieux à des effets non anticipés mais aussi à l'évolution des connaissances et ressources disponibles sur ce sujet, il est indispensable de paramétrer un suivi et des évolutions en fonction du contexte.

Ressources

[“L'effet rebond”, ressources du Mooc Impacts environnementaux du numérique \(INRIA\)](#)

[Les impacts environnementaux et sociétaux des données: un défi pour l'avenir](#)

Le cycle de vie de la donnée



Planification



OBJECTIFS VISÉS

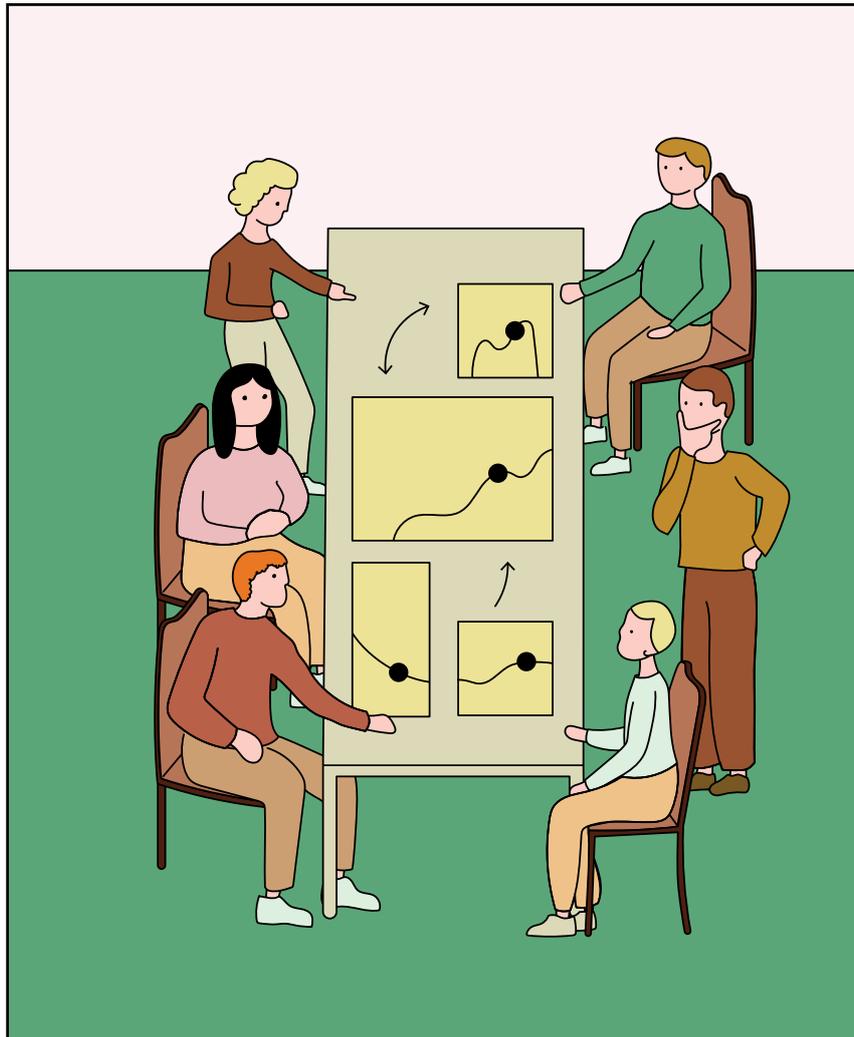
- définir une stratégie de gestion et de sobriété des données, en explicitant les modalités d'ouverture, de maintien, et d'accès aux données ouvertes
- garantir l'engagement des membres de l'organisation (SI, direction numérique) et de son écosystème (innovateurs, utilisateurs) à travers l'élaboration d'une charte et de sessions de sensibilisation
- mettre en place un groupe de pilotage d'une démarche visant à la sobriété des données et des services de données ouvertes chargé d'évaluer et suivre différents indicateurs d'impact environnementale de ses activités
- cibler les applications métiers, les usages et usagers prioritaires à accompagner dans la prise la production
- mobiliser les données ouvertes

PILOTES

- DPO (Délégué à la Protection des Données) direction numérique
- DSI (Direction des Systèmes d'Information) RSSI (Responsable de la Sécurité des Systèmes d'Information)
- service des archives
- service des achats

BP 1 - ENGAGER ET CONDUIRE UNE DÉMARCHE "DONNÉE OUVERTE ET RESPONSABLE"

BP 2 - PRIORISER LES CHOIX DE DONNÉES PUBLIÉES



BP1 - ENGAGER ET CONDUIRE UNE DÉMARCHE

“DONNÉE OUVERTE ET RESPONSABLE”

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Direction générale

INDICATEURS DE PILOTAGE

- Constitution et animation d'un groupe chargé de l'élaboration et suivi de la charte: oui/non
- % des collaborateurs de l'organisation ayant suivi une forme de sensibilisation aux enjeux de la charte
- Présence dans l'organisation d'une personne déléguée ou référente à la maîtrise des données ouvertes: oui/non

RETOUR D'EXPÉRIENCE

La [charte de la donnée de la métropole de Nantes](#), propose un chapitre dédié à la question des enjeux de sobriété des données, afin “d'agir avec sobriété et dans le respect des objectifs de la transition énergétique”. La métropole explicite aussi qu'il n'est pas souhaitable de collecter et stocker massivement des données dont l'utilisation n'est pas nécessaire.

RESSOURCES

- [Fiche de missions et tâches d'un.e responsable GreenIT](#) proposé par le CIGREF
- Exemple de [Charte Numérique Responsable](#) développée et diffusée par l'INR

Contexte :

Afin de coordonner et suivre les pratiques visant à réduire l'empreinte environnementale des services open data, il est recommandé de désigner un responsable (ou un groupe responsable).

Cette bonne pratique vise à préciser le cadre et les missions de ce référent mais aussi, afin de mieux acter et diffuser la démarche, de travailler à l'élaboration d'une charte de la donnée ouverte et maîtrisée.

Description de la bonne pratique :

1 - Identifier et nommer une personne responsable du suivi de la démarche

Recruter, nommer ou affecter une personne dédiée à la mise en place et à la coordination des actions relatives à l'évaluation d'impact et la maîtrise des données ouvertes et services associés, au sein de l'organisation. Dans les petites structures, un référent Numérique Responsable peut être nommé au sein de l'entité, en plus de ses autres tâches. Une possibilité supplémentaire est de partager cette responsabilité au sein des équipes.

Cette personne devra non seulement travailler avec la direction du numérique mais aussi de façon transverse dans l'organisation (RSE, service achats, les directions dites « métiers »...). Elle aura pour mission de :

- coordonner et suivre les initiatives de maîtrise d'impact des données en interne et en externe (auprès des prestataires, partenaires, utilisateurs...)
- développer un réseau de référents pour faire vivre et inscrire la démarche dans la durée
- identifier des mauvaises pratiques d'un service open data et des bases de données
- identifier et caractériser les impacts directs, indirects et des effets rebond de ces services
- mener une campagne de sensibilisation en interne et/ou vis à-vis des fournisseurs et des prestataires
- identifier et suivre un budget spécifique à l'animation de cette démarche

2 - Acter la démarche dans une charte ou un manifeste

Chaque organisation peut créer sa propre charte selon un processus participatif. Il est préférable que la charte soit modeste mais appliquée, plutôt que conséquente et sans réel engagement. Cette charte peut être révisée chaque année. Le processus participatif, et donc l'implication de chacun, est important pour l'application concrète de la charte. Élaborée et validée par les acteurs mobilisés dans les actions de production et d'utilisations de données ouvertes, une charte donnée ouverte et responsable vise :

- acter l'engagement de l'organisation, à impliquer d'abord ses collaborateurs, mais aussi ses partenaires et son écosystème, qu'il soit producteur ou utilisateur de données ouvertes

/ PLANIFICATION

- diffuser en interne de l'organisation des actions de sensibilisation. Le préalable à tout changement volontaire et engagé est la connaissance et la compréhension des enjeux

A titre d'exemple, la charte de la donnée ouverte et responsable pourrait faire figurer certains de ces principes d'actions:

- ouverture et transparence – accessibilité (sans authentification, design, médiation), licence (Etalab, ODBL)
- sobriété et maîtrise – granularité des données (temporelle, géographique), réduire la volumétrie et la duplication
- hébergement et souveraineté – évaluer et contrôler son hébergement (PUE, WUE, BEGES, localisation), critères d'achat responsables
- au service des usages et des transitions – des données tirées par les usages (priorité, capacitation, attentes et dynamique d'usages), qualité des données (standards, documentation des métadonnées, curation)

Il convient de corrélérer ce travail avec les autres initiatives animées ou en projet au sein de l'organisation.

Notamment celles liées au numérique responsable et l'écoconception numérique, ou encore à l'achat public responsable.

BP2 - PRIORISER LES CHOIX DE DONNÉES PUBLIÉES

PRIORITÉ

● ● ○ recommandée

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- Identification de critères, relais et de thèmes prioritaires pour prioriser l'ouverture des données.
- % de données ouvertes associées à un standard (Socle Commun des Données Locales ...)

RETOUR D'EXPÉRIENCE

Sur la base des intérêts présentés, voici quelques exemples de données ouvertes à fortes valeurs d'usages, publiées par les territoires et respectant le Socle Commun des Données Locales:

[Ville de Toulouse – Menus des cantines](#)
[Ville d'Antibes – Délibération du Conseil Municipal](#)

RESSOURCES

- [Carte de l'Observatoire OpenData](#)
- [Annuaire des plateformes territoriales open data](#)

Contexte :

Depuis octobre 2018, les collectivités territoriales de plus de 3500 habitants et employant plus de 50 agents (en équivalent temps plein) sont tenues de mettre en ligne:

- les documents qu'elles communiquent en application des procédures prévues par le CRPA, ainsi que leurs versions mises à jour
- les documents qui figurent dans le répertoire des informations publiques (RIP)
- les bases de données, mises à jour de façon régulière, qu'elles produisent ou qu'elles reçoivent et qui ne font pas l'objet d'une diffusion publique par ailleurs
- les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental
- les règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions, lorsque ces traitements fondent des décisions individuelles

Description de la bonne pratique :

Prioriser la publication des données qui représentent un intérêt particulier pour le territoire. Cette bonne pratique vise à prioriser les choix de données en prenant en compte le cadre légal d'ouverture des données, les standards disponibles, ainsi que l'intérêt pour l'organisation et son territoire.

Afin de limiter l'impact environnemental des données ouvertes et services associés, il convient ici de prioriser la publication de telle ou telle donnée, au regard de son:

- intérêt démocratique: commande publique, délibération, etc.
- intérêt économique: transport, équipements publics, énergie, commerce
- intérêt sanitaire: Covid-19, localisation défibrillateurs
- intérêt environnemental: espaces protégés, qualité de l'air, de l'eau, risque naturel
- intérêt social et culturel: agenda culturel
- intérêt politique local: un jeu de données qui concourt à un projet transverse prioritaire pour la collectivité, par exemple "Territoire Zéro Carbone" à La Rochelle

En complément, voici une liste de questions à se poser avant de collecter et publier un jeu de donnée:

1. Le jeu de données existe t-il déjà?

Identifier et associer les nouveaux jeux de données aux ressources déjà existantes. Les données pivots permettent précisément de rattacher de nouvelles données à d'autres jeux de données.

2. Le jeu de données rentre-t-il dans le cadre général d'ouverture des données tenues de rendre publiques ?

La CNIL nous rappelle dans [ce document](#) le cadre légal à l'ouverture des documents administratifs et données

spécifiques (délibérations, achats et subventions publiques, ...).

3. Le jeu de donnée est-il intégré dans le Socle Commun des Données Locales (SCDL) ?

Ce socle commun des données locales vise à créer une base commune de publication pour les collectivités locales et à normaliser la publication de ces données. Ces données sont soumises à des standards pour favoriser leur qualité et interopérabilité.

4. Quels sont les jeux de données publiés par les organisations ou collectivités de même niveau ?

Les compétences et périmètres d'actions étant spécifiques à chaque échelon, des territoires plus matures peuvent être sources d'inspiration sur les données à prioriser à l'ouverture.

5. Des besoins récurrents de jeux de données sont-ils déjà exprimés (en interne et externe de l'organisation) ?

Préférer une ouverture des données guidées par les usages plutôt que par l'offre permet de valoriser rapidement les données.

Production

—

OBJECTIFS VISÉS

- collecter et mettre à jour des données (à partir d'applications, de remontée de capteurs ou de saisies manuelles)
- s'aligner sur les standards et formats nécessaires à l'interopérabilité, l'exploitation et l'éditorialisation des jeux de données
- documenter et décrire précisément les métadonnées (données qui décrivent le contenu d'un jeu de données) afin d'explicitier le contexte de production, la nature du jeu de données, et de faciliter ainsi leur identification et réutilisation

PILOTES

- services producteurs de données

BP 3 - STANDARDISER LES JEUX DE DONNÉES

BP 4 - DOCUMENTER PRÉCISÉMENT LES MÉTADONNÉES

BP 5 - PROPOSER UNE GRANULARITÉ TEMPORELLE ET GÉOGRAPHIQUE ADAPTÉE



BP3 - STANDARDISER LES JEUX DE DONNÉES

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- % de données ouvertes publiées associés à un schéma de données

RETOUR D'EXPÉRIENCE

Pour garantir la réutilisation de ses données, et donc que l'usage fait soit plus important que le coût environnemental de publication, la ville de Pirae a fait le choix de ne publier que des données standardisées. [Ville de Pirae - Données ouvertes basés sur le SCDL](#).

Aussi, afin d'encourager les réutilisation, les données ouvertes accessibles sur data.gouv.fr proposent un filtrage par schémas..

RESSOURCES

Pour faciliter la consultation et récupération des jeux de données, il convient de les publier dans un format ouvert, standard, et non-proprétaire. Pour chaque grand type de données, il existe un ou plusieurs format standard, communément accepté et reconnu par la communauté d'utilisateurs et d'agents.

Contexte :

Afin de garantir leur qualité et leur interopérabilité, certains jeux de données font l'objet d'une standardisation. Les données standardisées sont choisies en fonction des demandes et des usages les plus fréquents.

La standardisation des données offrent un cadre méthodologique et technique à la publication. Elle vise à préciser quels sont les données essentielles, les formats préconisés (codification, unité, ...), la structure du jeu de donnée (ordre de présentation, données obligatoires, ..). L'utilisation de standard, comme proposé à travers le Socle Commun des Locales:

- invite les collectivités à prioriser l'ouverture des données territoriales particulièrement intéressantes dans le champ de la transparence ou de la création de services, et pouvant constituer un socle national
- aide les producteurs à ouvrir et à améliorer la qualité des données qu'ils publient.
- facilite l'interopérabilité des données et leurs exploitations par les réutilisateurs (agrégation, consolidation et traitements automatiques).

Description de la bonne pratique :

1- Voir et étudier les standards de données existants

Les schémas disponibles, en cours d'élaboration ou à proposer sont accessibles via schema.data.gouv.fr. Ce service aide les producteurs de données à connaître et comprendre les gabarits de données attendus.

La documentation des schémas précise les modèles de données :

- description des données essentielles
- format (chaîne de caractères, nombre entier, géolocalisation à x décimales, date...)
- champ obligatoire ou optionnel
- modalité de mise à jour
- etc.

2 - Associer un jeu de données à un schéma existant

Il est recommandé d'indiquer si un jeu de données correspond à un schéma depuis l'interface d'administration de data.gouv.fr. Le fait d'indiquer qu'une ressource respecte un schéma permet de bénéficier de vérifications de la qualité des données et d'indiquer aux réutilisateurs que vos données respectent un standard.

Certains producteurs choisissent de faire figurer dans le titre ou la description si le jeu de donnée est associé à un standard. Il serait aussi envisageable de le préciser dans les métadonnées, en intégrant dans la description un lien à la documentation du standard utilisé.

3 - Encadrer les prestataires pour garantir la compatibilité de jeux de données produits aux schémas existants

Lorsque les données produites dans un cadre de marché le titulaire peut être tenu de transmettre ces données à la collectivité dans les formats décrits dans le standard.

Les formats de transmission des données autres que celles mentionnées dans le référentiel national Socle Commun des Données Locales, seront transmises à la collectivité sous un format ouvert qui devra être décrit dans le mémoire technique.

4 - Participer à l'élaboration des standards

Les standards ne sont pas imposés par une «force» extérieure. C'est un référentiel collectif qui doit être enrichi et entretenu par tous les producteurs potentiels de données au profit des usagers finaux. Vous pouvez donc participer à l'élaboration de standards qui vous paraissent pertinents, ou demander des évolutions dans les standards existants pour les rendre plus aisés à respecter en fonction de vos contraintes de production.

BP4 - DOCUMENTER PRÉCISÉMENT LES MÉTADONNÉES

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- % de Jeux De Données (JDD) disponibles intégrant des indicateurs de sobriété dans les métadonnées
- Nombre de JDD ayant + de x% de champ N/A dans les métadonnées

RETOUR D'EXPÉRIENCE

Le portail data.gouv.fr expérimente un score de qualité des métadonnées pour chaque jeu de données publié. Ce score vise à inciter les producteurs à mieux renseigner les métadonnées et à orienter les utilisateurs dans le choix de données de qualité.

RESSOURCES

- [Documenter les données avant publication - Rédiger la fiche métadonnée associée au jeu de données et la tenir à jour - OpenDataFrance \(2018\)](#)
- [Guide de saisie des éléments de métadonnées INSPIRE—Recommandation du Conseil National de l'Information Géographique \(2013\)](#)
- [Guide de mise en oeuvre du schéma DCAT-AP rédigé par Pascal Romain du Département de la Gironde pour OpenDataFrance à partir du draft final de la spécification de la Commission Européenne \(2014\)](#)
- [Data Catalog Vocabulary \(DCAT\) - Recommandation W3C relative au vocabulaire des catalogues de données publiés sur le web \(2014\)](#)

Contexte :

Les métadonnées sont des informations structurées qui décrivent, localise et facilite la gestion d'un jeu de données, par exemple l'identification du producteur, la zone géographique concernée ou la date de mise à jour.

Elles permettent de comprendre exactement la nature des données qui sont mises à disposition. Grâce à une bonne compréhension des données disponibles, on peut donc choisir les données les plus intéressantes pour un usage donné et réduire les téléchargements inutiles.

Les métadonnées sont aussi une bonne source d'information pour mesurer la qualité des données et leur politique de publication (présence d'indicateur sur le respect d'un standard par exemple).

Enfin, les métadonnées semblent être un espace intéressant pour préciser des indicateurs de sobriété des données mises à disposition. Cette piste sera utilisée ultérieurement car dans l'état actuel, les métadonnées, calquées sur le référentiel DCAT, n'intègrent pas de champs spécifiques en lien avec la sobriété. On pourrait faire apparaître le nombre de chargement, de réutilisation ou le volume des ressources associées. Les métadonnées permettent de:

- réduire les barrières d'accès aux ressources pour une meilleure visibilité et une augmentation du potentiel de réutilisation des jeux de données
- mieux identifier les jeux de données disponibles et associés (par thèmes notamment)
- limiter les téléchargements inutiles grâce à une description fine des métadonnées

Description de la bonne pratique :

1 - Les données essentielles à figurer dans les métadonnées

Décrire avec attention et précision les métadonnées de chaque jeu de données. Les informations généralement attendues pour décrire un jeu de données sont les suivantes:

- Titre du jeu de données
- Description libre de l'objet et du contenu de la donnée
- Thème du jeu de données
- Nom de la structure qui diffuse la donnée
- Nom de la structure qui produit la donnée
- Nom de la structure qui gère la donnée
- Couverture spatiale sur lequel s'appliquent les données
- Début/fin de la plage temporelle couverte par les données
- Fin de la Plage temporelle couverte par les données
- Date de la première publication
- Fréquence de la mise à jour
- Date de la dernière mise à jour publiée
- Mots-clés permettant des recherches libres
- Licence appliquée sur le jeu de données
- Liste des formats où sont publiées les données
- Code de la projection géographique quand cela s'applique
- Langue du jeu de données
- Liens vers les ressources accessibles

2 - Utilisation des normes existantes

Appliquer de préférence un format de métadonnées normalisé. Pour garantir cette homogénéisation des métadonnées, il est important de s'appuyer sur des formats reconnus pour s'assurer de leur pertinence (les données nécessaires) et de leur format (la façon dont on les a structurés et codifiés).

Les formats standards permettent la collecte et la recherche des données et autorisent un traitement automatique des métadonnées. Les métadonnées appliquées aux données ouvertes en France sont une mise en application opérationnelle du schéma international DCAT.

Dans de nombreux cas, c'est le portail qui héberge les données qui propose d'adopter tel ou tel format pour les métadonnées. Ceux-ci sont généralement en conformité avec les standards internationaux. Les standards recommandés sont INSPIRE pour les données géographiques, DCAT et ses déclinaisons pour tout type de données ouvertes.

Bien qu'il n'existe pas un modèle de métadonnées open data unifié, OpenDataFrance a proposé un standard dans le jeu de données Catalogue du Socle Commun des Données Locales.

3 - Développer et intégrer de nouveaux indicateurs dans les métadonnées

Intégrer dans les métadonnées des informations spécifiques à la gestion responsable des données:

- indicateurs de qualité – conforme à un standard existant (O/N ou tag + lien du standard), granularité et périmètre du jeu de données
- indicateurs de sobriété – volume des ressources, compression utilisée, donnée "froide"
- indicateurs de gestion – formats des datasets, politique d'archivage, durée de vie de la donnée
- indicateur de souveraineté (optionnel) – hébergement de la donnée, protection des données

A noter que la curation et la gestion des métadonnées doit être opérée par le producteur des données lui-même.

BP5 - PROPOSER UNE GRANULARITÉ TEMPORELLE ET GÉOGRAPHIQUE ADAPTÉE

PRIORITÉ

● ● ○ recommandée

MISE EN OEUVRE

● ● ● difficile

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- % de jeux de données proposant une granularité temporelle ou géographique référente à la maîtrise des données ouvertes: oui/non

RETOUR D'EXPÉRIENCE

La plateforme opendatasoft propose de récupérer les données disponibles en filtrant différents champs: couverture spatiale ou temporelle, producteurs, thématiques...

Par exemple, ici, le CD34 (Data Hérault) propose d'isoler les aménagements cyclables par filtres: communes, vitesse maximale pour le trafic adjacent, type d'aménagements...

RESSOURCES

- API Adresse (Base Adresse Nationale - BAN) - Documentation de cas pratiques d'utilisations des API géographiques par Etalab (2022)
- Les tuiles vecteur openmaptiles.geo.data.gov/fr
- Cette API permet de mettre à disposition des tuiles vectorielles qui sont affichables sur des cartes géographiques interactives. Elles servent principalement à afficher des fonds de plan mais aussi les contours cadastrales et les limites administratives en France. Cela permet de s'affranchir d'APIs cartographiques comme Google Maps.

Contexte :

La production et la diffusion de jeux de données sobres visent à limiter les flux et le nombre de téléchargements de données non utilisées. Offrir un mode de récupération des jeux de données permettant de cibler une temporalité et/ou une cellule administrative spécifique permet notamment :

- de réduire les flux de données transmises
- de limiter la volumétrie des données stockées chez l'utilisateur
- de mieux répondre aux besoins des usagers
- d'anticiper la gestion de l'historicité des données

Description de la bonne pratique :

1 - Lors de la production des données

Décrire avec précision la granularité temporelle et géographique des données contenues dans les jeux de données. Pour chaque jeu de donnée, il convient de préciser quelle est la plus petite échelle géographique dans les données présentées, par exemple : la zone IRIS, la commune, le canton, etc.

De même, on précisera la granularité temporelle, c'est à dire l'échelle, pour les données essentielles : heure ou jour (cas de la pollution de l'air), année (cas des budgets), etc. Ces informations sont différentes de la notion de couverture:

- couverture géographique : périmètre géographique qui correspond à l'emprise maximale géographique des données précisées, par exemple la «ville de Rennes» (dans le monde géographique, on précise parfois les coordonnées des limites du rectangle qui contient les données (un point en haut à droite, et un point en bas à gauche)
- couverture temporelle: période temporelle contenant les données présentées, par exemple – "2021" pour un jeu de données sur les budgets de telle commune, élections "2018 à 2022" pour les résultats des élections des cinq dernières élections de telle commune

Ces champs indiquent le choix qui est fait dans le découpage et la présentation des données, ils permettent aux réutilisateurs de cibler précisément les données à récupérer et de limiter les données transférées (volume réduit). Dans certains cas, le portail de données ouvertes peut proposer un service de création de filtres ou de sélection.

2 - Lors de la récupération des données

En tant qu'utilisateur d'un portail de données ouvertes, il est courant de rechercher des données spécifiques publiées dans un jeu de données. En appliquant certains filtres, il est possible d'isoler seulement les données nécessaires et attendues plutôt que de récupérer la totalité du jeu de données.

3 - Export de données géolocalisées

GéoDataMine et DataClic sont des outils simples d'extraction thématique de données OpenStreetMap ou de bases nationales. Ils permettent de :

- choisir un territoire (commune, EPCI, département)
- choisir une thématique (parking à vélo, défibrillateurs, enseignement, entreprises, élections...)
- récupérer les données issues d'OpenStreetMap ou d'une base nationale sur ce thème et sur ce territoire (formats CSV, GeoJSON, Shapefile)

Analyse

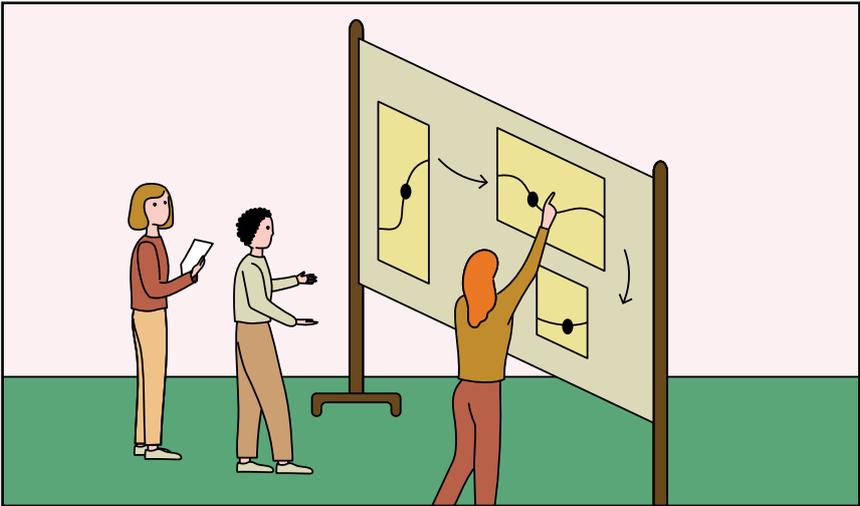


OBJECTIFS VISÉS

- contrôler la qualité des données produites pour s'assurer de leur fiabilité sur la forme et sur le fond (modifications, des traitements, des corrections ou des enrichissements)
- explorer les données collectées et produites afin d'identifier des relations entre différentes variables et échelles
- interpréter et rendre compte d'indicateurs de pressions, d'états ou de réponses utiles à la conduite et au suivi de politiques publiques
- des actions visant à limiter et écoconcevoir les datavisualisations pourraient réduire l'impact environnemental

PILOTES

- services utilisateurs de données



Publication



OBJECTIFS VISÉS

- limiter l'impact environnemental des services OpenData
- la découvrabilité et la réutilisation par le référencement des données sans multiplier les duplications
- l'export et l'import automatisés et ciblés de jeux de données dans des applications métiers spécifiques grâce à des API

PILOTES

- services producteurs et utilisateurs de données

BP 6 - RÉDUIRE LE VOLUME UNITAIRE DES JEUX DE DONNÉES

BP 7 - PROPOSER UN ACCÈS AUX DONNÉES PAR API

BP 8 - FACILITER LA DÉCOUVRABILITÉ DES JEUX DE DONNÉES



BP6 - RÉDUIRE LE VOLUME UNITAIRE DES JEUX DE DONNÉES

PRIORITÉ

● ● ● recommandée

MISE EN OEUVRE

● ● ● difficile

PILOTES

Référent aux données ouvertes et responsables, DSI

INDICATEURS DE PILOTAGE

- Volume moyen des fichiers disponibles en Mo
- % de jeux de données supérieur à 240 Mo

RETOUR D'EXPÉRIENCE

Accès aux résultats des élections à Grand Poitiers: La manière de gérer les données des élections municipales à Grand Poitiers est remarquable. Un jeu de données indépendant pour chaque élection, dont aucun ne dépasse 40 Ko. A l'inverse, une autre collectivité publie par exemple le résultats de toutes les élections depuis 1992 dans un même fichier. Le fichier .csv résultant est de 42 Mo, soit x1000 plus volumineux que celui d'une seule élection à Poitiers.

Volumétrie des jeux de données limitée par les éditeurs de portails: Certaines plateformes comme OpenDataSoft limite la taille maximum d'importation d'un jeu de données à 240 Mo. Au-delà, il convient soit de compresser ou de fragmenter le jeu de données.

RESSOURCES

- [Le Big Bang du Big Data](#)

Contexte :

Les premières externalités négatives de la donnée sont liées à son stockage et sa diffusion qui nécessitent une infrastructure matérielle (réseau, et data centers), mais aussi des terminaux utilisateurs. Cela se traduit naturellement par de la consommation de matière première (minerais, eau, beaucoup d'eau) pour produire ces équipements, et de la consommation électrique.

Non seulement la production de données est colossale mais elle croît à une vitesse vertigineuse en raison des usages, des multiples sources (IoT) et leur distribution/duplication via internet.

Nous le savons, l'OpenData propose généralement des jeux de données au volume limité, sans commune mesure à certaines ressources ou services mobilisant des données en temps réel, voire des données cartographiques.

En 2021, le Shift project estime la part du numérique à 12% de la consommation totale d'électricité et 3 à 4 % des émissions de gaz à effet de serre (GES) dans le monde. Par ailleurs, la production de donnée au niveau mondial est exponentielle. Portée par une explosion des usages, en 2020, les données générées, copiées et consommées en ligne est estimée à plus de 60 Zettabytes (1 zetta représente environ 1 milliard de terabytes).

Bien que ce volume soit essentiellement porté par une croissance forte du contenu vidéo et temps réel, et que la volumétrie seule n'est pas un indicateur suffisant d'impact du numérique, une attention particulière peut être portée sur la volumétrie des données produites, utilisées par les métiers, exposées et archivées.

Description de la bonne pratique :

Bonnes pratiques ayant un impact important sur la réduction de l'empreinte environnementale:

1 - Privilégier le format de fichier plus sobre

Les formats de fichiers peuvent être plus ou moins «gourmands» : par exemple, il peut y avoir un rapport de 1 à 10 entre le volume d'un fichier au format .csv et les mêmes données dans un format .xls

2 - Réduire le nombre de ressources

Limiter le nombre de formats proposés si cela n'est pas indispensable. Il est bien entendu possible de présenter plusieurs formats pour un jeu de données. Cela facilite le traitement du point de vue du réutilisateur. Mais il faut avoir conscience que chaque duplication avec un format particulier augmente le volume de données stockées

La co-existence fréquente de multiples formats (csv, json, xls, shp, ...) pour un même jeu de données doit poser la question de la pertinence de cette multiplication. Sans la bannir, cette pratique doit être mise en oeuvre si elle est vraiment indispensable.

3 - Proposer la récupération des informations sous forme d'API

Proposer systématiquement un accès sous format d'API dans le cas des données volumineuses ou à changement d'état fréquent afin d'effectuer des requêtes pour filtrer/trier les données à la source.

Lorsque le fichier est trop volumineux, il vaut mieux accéder directement à une partie des données à travers les critères d'accès que proposera un service d'accès informatique ou API.

De même pour les données changeant d'état fréquemment (par exemple; la disponibilité en temps réel des parkings), il sera plus pertinent de ne charger que les enregistrements récents ou les derniers changements plutôt que la base entière.

On ne réduit pas le fichier source mais uniquement les données transmises.

Bonnes pratiques ayant un impact modéré sur la réduction de l'empreinte environnemental:

4 - formater le fichier pour éviter les redondances

Choisir le format de présentation des données le plus adapté. On trouve parfois plus pratique de publier les données au format tabulaire alors que certains jeux de données sont plus légers dans un format type json. Vouloir les rendre tabulaires multiplie énormément les redondances dans les chaînes de caractère communes.

5- réduire les informations aux données essentielles (filtre à la production)

Supprimer les colonnes redondantes (doublet ou déductibles) et réduire la longueur des champs dans la mesure du possible. Par exemple, si un jeu de données précise le code INSEE d'une commune, inutile de fournir son nom, son département ou sa région. Ces informations peuvent être obtenues par croisement avec un référentiel.

6 - fragmenter le fichier pour permettre des téléchargements partiels à l'utilisateur

Découper les ressources avec une granularité et une couverture géographique et temporelle adaptée. Par exemple, un fichier des résultats des 10 dernières élections peut être découpé en 10 fichiers, un pour chaque élection.

7 - Compresser le fichier pour optimiser les téléchargements

Il est possible de proposer au téléchargement des fichiers compressés lorsque le volume et l'efficacité de la compression le justifie. Certains portails offrent des fonctions de compression lors des transferts (exemple : opendatasoft).

BP7 - PROPOSER UN ACCÈS AUX DONNÉES PAR API

PRIORITÉ

● ○ ○ pour aller plus loin

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- % de données ouvertes disponibles par API

RETOUR D'EXPÉRIENCE

Contrôler les flux d'API – ADEME: Pour protéger l'infrastructure de publication de données de l'Ademe, les appels sont limités par quelques règles simples : Un utilisateur anonyme ne peut pas effectuer plus de 100 requêtes par intervalle de 5 secondes. Sa vitesse de téléchargement totale est limitée à 2 MB/s pour les contenus statiques (fichiers de données, pièces jointes, etc.) et à 500 kB/s pour les autres appels.

RESSOURCES

- Des standards tel que OpenAPI peuvent aider et s'intégrer très rapidement sur api.gouv.fr
- Si vous souhaitez mettre à disposition et référencer votre API, vous pouvez l'enregistrer sur ce portail : <https://api.gouv.fr/nouvelle-api>

Contexte :

L'interopérabilité pour faciliter l'accès et la réutilisation de données:

Le Web est une machine à développer des services. Basés sur une architecture dite «REST», de nombreux services peuvent, selon leur configuration, consulter et mettre à jour facilement des données ouvertes.

Afin de faciliter l'accès et la réutilisation de données, la mise en place d'interfaces de programmation applicative (API) vise à mettre à disposition des données dans un format facilement utilisable. Plutôt que de dupliquer et héberger soi-même des jeux de données, les API facilitent les récupérations et actualisations. Ainsi, chaque donnée mise à jour est automatiquement actualisée dans l'ensemble des services les mobilisant par API.

Bénéfices d'accès aux données par API :

- ne transférer que les données nécessaires au bon moment (filtre des données à la source)
- ne pas multiplier les duplications et réduire les besoins de stockage
- faciliter la récupération et la valorisation et l'usage des données ouvertes

Description de la bonne pratique :

Architecture

Certains éditeurs de portails OpenData proposent de manière systématique le déploiement d'API. Cette fonctionnalité facilite de fait l'interopérabilité et la récupération de jeux de données.

Interopérabilité

Pour garantir ce critère d'interopérabilité des données et des services mais aussi harmoniser et faciliter la récupération de données, il est recommandé d'utiliser le principe d'architecture REST et le principe d'encodage JSON. D'autres formats pourraient être à encourager notamment pour les données dites chaudes, temps réels, comme le format BSON ou encore CBOR.

Critères d'achat

Dans certains cas, il peut être pertinent d'intégrer des critères de mises à disposition d'API et modalités d'accès, dans les cahiers des charges qui encadrent les prestataires (assurer le développement d'API, garantir une infrastructure propice à la mise à disposition d'API)

Documentation

Pour chaque API proposée, il convient d'explicitier:

- une documentation technique présentant les modalités d'interrogation, rythme d'actualisation, formats ou modalités de récupération de la donnée – des standards tel que OpenAPI peuvent aider

- les conditions générales d'utilisation qui précisent le processus de demande d'accès, la durée ou encore les critères d'éligibilité des réutilisations

API existantes

Le service api.gouv.fr propose de nombreuses données ouvertes et la documentation nécessaire pour aller les récupérer et les intégrer aux services et portails OpenData. Voici quelques exemples de bases de données accessibles par API spécifiques :

- géorisques naturels et technologiques des territoires
- piézométrie (niveau des nappes d'eau souterraine)
- l'ensemble des adresses du territoire français
- les Diagnostic de Performance Energétique des logements particuliers (DPE)

L'accès par API ne peut être exclusif

La CADA a récemment statué que « la consultation sur Internet de documents librement communicables ne saurait être subordonnée à une procédure de demande d'accès impliquant une autorisation préalable de l'administration ». Autrement dit, bien qu'un accès par API peut-être pratique et facilitant, il ne doit être exclusif et opérer un accès restreint à des données publiques.

BP8 - FACILITER LA DÉCOUVRABILITÉ DES JEUX DE DONNÉES

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Référent aux données ouvertes et responsables

INDICATEURS DE PILOTAGE

- Nombre (ou %) de jeux de données récupérés par API
- Nombre (ou %) de jeux de données référencés sur plus d'un portail

RETOUR D'EXPÉRIENCE

La ville de Castelnaudary publie ses données ouvertes directement sur le portail data.gouv.fr. L'exposition et la valorisation des données se fait localement, sur le site officiel de la ville :

Le portail OpenData de la Région Occitanie référence l'ensemble des données produites par le Conseil Départemental de l'Hérault et facilite ainsi leur découvrabilité et le croisement de cas d'usages.

RESSOURCES

- [Demander à data.gouv.fr de moissonner votre site](https://data.gouv.fr)

Contexte :

Les données dupliquées et redondantes consomment des ressources matérielles (les disques de stockage) et de façon récurrente de l'énergie (pour la redondance). Cette pratique vise à faciliter la découvrabilité des données plutôt que leur duplication afin de libérer de l'espace sur le stockage et d'optimiser l'utilisation de données ouvertes.

Plus la donnée est référencée, plus elle est susceptible d'être identifiée et donc utilisée. De fait son coût, technologique et environnemental, peut-être justifié par les usages qu'offre la donnée. Développer le référencement permet pour le producteur de données de :

- faciliter et mieux suivre les réutilisations des données
- permettre des agrégations territoriales ou thématiques
- faire des économies d'échelle (technique et financières) par la mutualisation

Description de la bonne pratique :

1 - Éviter la duplication des données

Afin de réduire l'impact environnemental et le coût de stockage de données ouvertes, il convient de limiter leur duplication et de privilégier le moissonnage des données. Une première étape peut s'intéresser à identifier les redondances des jeux de données volumineux et stratégiques : où les données sont publiées, dupliquées ...

2 - Identifier les lieux de publications de données

Des plateformes de données peuvent référencer et moissonner les (méta)données entre portails. Plutôt que de dupliquer un jeu de données, il est donc préférable d'optimiser son référencement en indexant le jeu de données sur d'autres portails. Par exemple :

- plateforme nationale ([data.gouv](https://data.gouv.fr))
- plateformes territoriales (échelon régional, départemental...)
- plateforme thématique (transport.data.gouv.fr, data.tourisme.gouv.fr, [geoportail...](https://geoportail.fr))

3 - Demander à data.gouv.fr de moissonner votre portail

Un moissonneur permet d'importer toutes les données d'un portail d'open data. Vous pouvez aussi demander au moissonneur de n'importer que certains jeux de données, au moyen de filtres.

4 - Maintenir l'accès et la qualité des données produites

Pour maintenir à jour les données référencées et limiter le risque de problème de version des jeux de données sur d'autres plateformes, privilégier un accès aux données par API. Voir la BP7 - Proposer un accès aux données par API.

Conservation

—

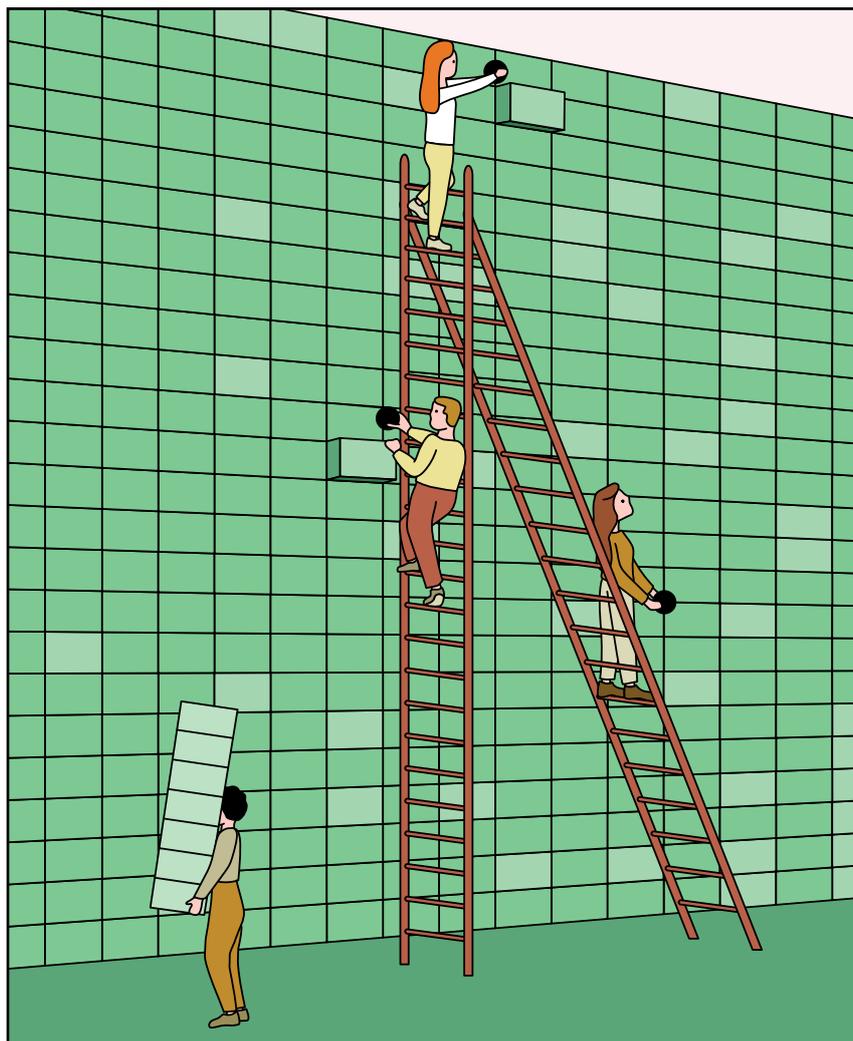
OBJECTIFS VISÉS

- définir et animer une stratégie d'archivage des données
- identifier quels sont les critères de suppression, d'indexage, et modalités d'accès à des données jugées comme obsolètes

PILOTES

- services des archives

BP 9 - METTRE EN PLACE UNE POLITIQUE D'ARCHIVAGE



BP9 - METTRE EN PLACE UNE POLITIQUE D'ARCHIVAGE

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Services des Archives

INDICATEURS DE PILOTAGE

- Volume de données archivées par an
- Nombre de références de données archivées
- Volume de données disponible directement via le portail Open Data
- Nombre de jeux de données travaillés avec archivistes

RETOUR D'EXPÉRIENCE

Mise en œuvre: La ville d'Antibes en lien étroit avec le service d'archive, démarre une réflexion et expérimente l'archivage de jeux de données très spécifiques (exemple, centre de vaccination), aujourd'hui dé-publié, mais pourtant caractérisé comme stratégique.

Identification: Le Ministère de l'Enseignement, de la Recherche et de l'Innovation a par exemple labellisé dans le titre le caractère d'archive d'un jeu de données "Archives - Appels à projets ANR - Projets retenus et participants identifiés".

RESSOURCES

- Le portail [opendataArchives](#) propose d'archiver et d'historiser l'opendata français. Il héberge aujourd'hui environ 50 To de données, soit plus de 70000 jeux de données issus de 300 portails.
- OAIS (Open Archival Information System, ISO 14721)
- Norme AFNOR NF Z42-013 présente les exigences opérationnelles et techniques à mettre en œuvre au sein du système d'archivage électronique

Contexte :

La conservation des données publiques par l'archivage vise 4 objectifs : rendre les données et documents produits accessibles, en garantir leur intégrité, leur intelligibilité via leur documentation et enfin leur lisibilité.

Il demeure une réalité organisationnelle où tout ne peut cependant être conservé en l'état dans les portails OpenData. Les services archives disposent de méthodes de suivi, de documentation et d'outils de gestion pour traiter les données publiques et leur conservation. La question de l'archivage des données ouvertes en open data n'est pas actuellement bien traitée (il existe même des activistes éthiques qui archivent l'open data de l'état). Il est pourtant légitimes de conserver des données qui ont été publiées, probablement exploitées, indépendamment de leurs sources (applications du système d'information) qui suivent une politique d'archivage propre. Et à ce jour, le problème de la conservation des données temps réel est rarement exploré (volumétrie, format, opportunité).

Cette pratique vise à développer une réflexion autour d'une stratégie d'archivage des données ouvertes au sein des organisations, afin de limiter l'impact environnemental du stockage de données ouvertes. On fera notamment la distinction entre données «vivantes», c'est à dire à jour et applicables, et les données «historiques», qui ne sont plus entretenus ni nécessairement à jour, mais dont l'accès doit être maintenu à de fins de consultation (suivi des politiques, comparaison, patrimoine informationnel).

Description de la bonne pratique :

1 - Animer et acter la démarche

Ce collectif aura pour tâche de définir la stratégie d'archivage des données ouvertes : qui décide de garder/supprimer/archiver sur temps long... Quelles sont les règles de nommage et de versioning ? Comment identifier, contrôler et à quelles conditions supprimer des données obsolètes stockées en masse ? Quels sont les moyens d'accès à ces données archivées ?

Ce groupe doit être animé par les membres du service des archives, et pourrait impliquer le référent aux données ouvertes, le délégué à la protection des données ou encore du responsable de la sécurité des systèmes d'information, de la personne responsable de l'accès aux documents administratifs, et des producteurs de données.

2 - Définir les conditions et critères d'archivage

Il est important de définir clairement pour chaque jeu de données, ses conditions d'archivage :

- quoi archiver – dans de nombreux cas, conserver un échantillonnage des données produites suffit à développer des modèles et analyses comparatives
- mécanismes d'archivage: durée d'utilité administrative

des données avant archivage temporaire et enfin pérenne avec accès sur demande

- formats ouverts et pérennes des données archivées, infrastructure sécurisée
- création et maintien d'un index des données ouvertes archivées
- modalités d'accès aux données archivées
- règles d'administration définissant les modalités d'accès aux archives

3 - Mieux identifier les données archivées

Séparer les données «vivantes», c'est à dire à jour, applicables ou pertinentes, des données «anciennes», éventuellement utiles à des fins de comparaison (suivi temporel) ou de recherche de version. La séparation évite de trouver ou utiliser des données inexploitablees ou avariées; elle permet aussi une gestion d'archivage différente.

Il peut être utile pour mieux identifier les données d'archives, de le labelliser dans le titre ou la description du jeu de données. Un attribut précisant le caractère d'archive d'un jeu de données peut aussi être précisé dans les métadonnées.

4 - Favoriser l'échantillonnage

Plutôt que de conserver l'ensemble d'un jeu de donnée, dans certains cas, en conserver un échantillon peut être suffisant pour limiter la charge de stockage tout en garantissant l'exploitation de modèles et d'analyses.

Exploitation

—

OBJECTIFS VISÉS

- critères environnementaux d'hébergement des données et services alignés au Code de conduite européen pour les centres de données
- pistes d'actions de mutualisation de services hébergées
- critères d'achat public responsable

PILOTES

- les DSI et services achats doivent être mobilisés dans cette étape afin de garantir le déploiement, le maintien des services et le suivi de leurs impacts

BP 10 - CONTRÔLER L'HÉBERGEMENT DES DONNÉES



BP10 – CONTRÔLER L'HÉBERGEMENT DES DONNÉES

PRIORITÉ

● ● ● prioritaire

MISE EN OEUVRE

● ● ○ moyenne

PILOTES

Responsable DSI, service des achats, RSSI, partenaires ou prestataires

INDICATEURS DE PILOTAGE

- Hébergement signataire du Code de Conduite européen des centres de données: oui/non
- % / listes des critères du code du Code de Conduite européen sur les Datacentres appliqués

RETOUR D'EXPÉRIENCE

La Région Bretagne travaille au sein d'un COPIL sobriété des infrastructures du système d'information à l'élaboration et au monitoring d'indicateurs qui permettent d'alimenter son BEGES numérique. Des actions cibles viseront aussi à l'optimisation des services hébergés et à la recherche de gisements applicatifs gourmands (plus forte empreinte ou plus forte empreinte/ utilisateurs).

RESSOURCES

- [Guide des bonnes pratiques du code européen sur les data centres](#) – Ecoinfo

Contexte :

Les impacts environnementaux des centres de données sont multiples, et dus en particulier à la consommation d'électricité pour l'alimentation et le refroidissement des équipements informatiques, à la fabrication de ces mêmes équipements, à l'artificialisation des sols pour l'installation du centre de données et à la consommation d'eau.

La Commission européenne se fixe notamment l'objectif de parvenir à des centres de données neutres sur le plan climatique d'ici à 2030 et annonce l'adoption de mesures en matière d'efficacité énergétique et d'économie circulaire pour les réseaux, les centres de données et les équipements, ainsi qu'un travail sur les marchés publics durables.

Le Code de conduite européen pour les centres de données vise à encadrer l'hébergement (SaaS ou On-premise) à travers différents indicateurs.

Description de la bonne pratique :

1 - Demander des éléments de transparence à son hébergeur

Contrôler l'hébergement de données confiées à des opérateurs externes (Cf Code de conduite européen pour les datacenters). Afin de réduire en valeur absolue l'empreinte environnementale des data centres, il peut être nécessaire de demander à son hébergeur des éléments de transparence sur:

- la localisation géographique – cette localisation va avoir une influence sur l'empreinte environnementale via le mix électrique du territoire concerné
- le taux de disponibilité – détermine le niveau de redondance des équipements de l'infrastructure
- la superficie (en m²)
- la densité électrique (kW/baie, kVA/baie)

2 - Prendre en compte des critères environnementaux lors du choix d'un prestataire d'hébergement

- adhésion au Code de conduite européen pour les datacentres (oui/non) ?
- gaz à effet de serre – utilisation d'équipements issus du réemploi (oui/non ou en %) et empreinte de la fabrication et de la fin de vie des équipements
- énergie – utilisation d'énergie électrique d'origine renouvelable (oui/non ou en %)
- éco-efficacité du service hébergé – kWh/Go
- power usage effectiveness – PUE < 1,2 (par exemple)
- eau – mesure et réduction de la consommation d'eau (WUE: water usage effectiveness – oui/non) et quantification (WUE < 1 L par kWh (par exemple))
- infrastructure – récupération de la chaleur fatale produite par les serveurs (oui/non), écoconception du bâtiment (oui/non), écoconception des serveurs (oui/non)

3 - Définir une stratégie d'hébergement responsable

Des décisions liées à des enjeux de gouvernance peuvent aider à limiter l'empreinte environnementale des services hébergés:

- définir une stratégie de mutualisation des ressources de stockage avec d'autres territoires/institutions ou services
- intégrer des critères d'achats responsables respectant les critères du Code de Conduite européen des centres de données
- définir un process efficace de scalabilité, augmenter ou réduire les ressources de stockage nécessaires selon les besoins réels
- s'aligner à un plan de souveraineté des données, en refusant par exemple l'utilisation de services hébergés ou monétisés hors Europe (GAFAM...), et privilégier les hébergements recommandés SecNumCloud (voir [ressource ANSSI](#))

POSTFACE

OpenData France est depuis 2013 un acteur français incontournable de la promotion de l'Opendata et de l'accompagnement des collectivités locales dans une démarche d'ouverture des données publiques. Ce référentiel Green Data construit après concertation des collectivités membres d'OpenDataFrance est un outil pratique pour mettre en œuvre une démarche d'ouverture des données tout en intégrant la dimension écoresponsable.

Dans une démarche de co-construction et d'ouverture des données publiques, les échanges entre Opendata France et la MiNumEco permettent d'enrichir nos pratiques et de permettre la transformation des organisations publiques françaises sur l'ensemble des champs d'amélioration des pratiques et usages du numérique.

La MiNumEco est la Mission interministérielle numérique écoresponsable, co-piloté par la Direction Interministérielle du Numérique (Services de la Première Ministre) et le Ministère de la Transition Écologique.

Depuis 2020, elle accompagne les administrations françaises dans leur transformation écologique, en fixant des objectifs ambitieux, en formant les agents, en produisant les outils nécessaires, et en accompagnant les services publics dans la réduction de l'empreinte environnementale du numérique public.

Tous ses outils et guides sont construits dans un souci à la fois d'exemplarité de l'État et d'ouverture à tous les publics : entreprise, citoyen, association. Une démarche et une pratique de la consultation, tant de l'écosystème, que des citoyens. Dans le cadre de cette démarche que nous partageons avec Opendata France, nous partageons les outils et référentiels pour l'administration, en Licence ouverte, comme le Guide de bonnes pratiques numérique responsable pour les organisations ou bien le Référentiel général d'écoconception de service numérique (RGESN).

Les outils de la MiNumEco sont à articuler avec ce référentiel de GreenData proposé par OpenData France, afin de promouvoir des pratiques numériques toujours plus responsables dans nos administrations, nos collectivités et toutes nos organisations.

*Virginie Rozière,
chefe du pôle conseil DINUM et responsable de la mission
MiNumEco.*

