



MINISTÈRES
TRANSITION ÉCOLOGIQUE
COHÉSION DES TERRITOIRES
TRANSITION ÉNERGÉTIQUE
MER

Liberté
Égalité
Fraternité



Feuille de route ministérielle de la donnée,
des algorithmes et des codes sources

Facilitateur de vos transitions
#CMVRH



" Explorer des jeux de données du pôle ministériel "
2ème webinaire d'un cycle sur la culture de la donnée
Vendredi 14 avril 2023 de 9h30 à 12h00

Introduction
Intervenante : Hélène BEGON - CGDD/SRI/Ecolab
Auteure de : La transformation numérique des administrations, La documentation française, août 2021
Support préparé avec les collègues de l'équipe data-IA de l'Ecolab du Commissariat général au développement durable



Enrichissements d'une donnée :

- Donnée brute, utile à son seul créateur (ou stockée pour d'autres modèles de traitement, des séries temporelles), surtout si elle est non numérique
(Encore que les chercheurs peuvent la préférer car elle n'est pas polluée d'interprétations, ou contrainte par un modèle conceptuel)
- Assortie de **métadonnées** (date, lieu, auteur, géoréférencement, licence de réutilisation, etc.)
- **De qualité** : récente, exhaustive, comparable dans le temps et homogène dans l'espace, **croisée avec des « données de référence »** (ex : base adresses nationale, code de la commune) etc. ; *il y a de nombreux critères de qualité de la donnée et ils dépendent des usages faits de la donnée*
- Inscrite dans un **schéma** partagé de données qui précise les formats des données, les métadonnées à associer, les valeurs possibles de champs, ...
- **Ouverte** en ligne, facile à trouver (inscrite dans des catalogues) et à comprendre, purgée des risques juridiques (donnée personnelle, droits de propriété, ...) - « Toute mise à disposition (...) sous forme électronique (...) se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé » (L.300-4 CRPA)
 - Ou au contraire **protégée**, cryptée... anonymisée, pseudonymisée...
- **Partagée** au sein d'un « Commun numérique » (communauté d'acteurs qui l'enrichissent collectivement) ou dans un environnement sécurisé
- **Projetée** sous forme d'animation vidéo, de data-visualisation, de cartes, de podcast, de publication...
- **Organisée** dans une base de données, un système d'information, une plateforme internet gérés en lien avec les besoins des utilisateurs
- Assortie d'**outils** pour la traiter, enrichir l'information qu'elle contient potentiellement : moteur de recherche, croisements de données, visualisations (schémas), visualisations multi-couches (cartes), algorithmes aidant à la prise de décision, puissance de calcul associée à l'outil, ...
- Largement et **massivement** accessible : API, croisements automatiques (type « Dites-le nous une fois »), haut débit de téléchargement...
- Pouvant servir à l'entraînement d'outils d'**intelligence artificielle**, et encore mieux si de la donnée est partagée entre de nombreux acteurs (espaces communs thématiques de données : ex *Health Data Hub*)
- Gérée et utilisée de façon à minimiser ses effets **environnementaux**
- Pensée dans un **cycle de vie** complet, y compris sa destruction le cas échéant.

Raisons d'être des outils qui permettent d'agréger et de donner accès aux jeux de données : systèmes de gestion de bases de données, systèmes d'information, plateformes internet, ...

Ceci se brouillant avec les outils qui extraient de la donnée à partir de sources dispersées, ou permettent de stocker ensemble des données non organisées : traitement du langage naturel, reconnaissance automatique d'image ou de mouvement, « lacs de données », « data scraping » automatisé sur internet, ...

Les données peuvent se présenter sous une forme plus ou moins organisée



Rapport d'activité
relatif aux autorités en
charge du cas par cas
des projets en 2021

Jun 2022

Donnée isolée, donnée brute, donnée non structurée

Jeu de données, collection de données (dataset en anglais)

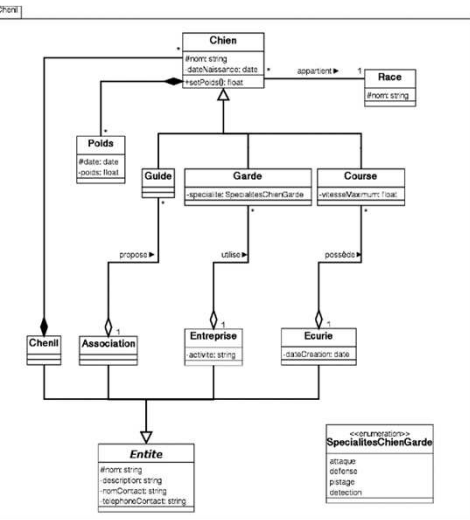
Ensemble structuré de données quelconques avec un thème commun, accessible en un ou plusieurs fichiers.

Toutes les données d'un jeu de données peuvent donc être décrites avec une majorité de métadonnées communes.

Il y avait 44 675 jeux de données sur le site de données publiques data.gouv.fr le jour où cette diapositive a été écrite

Tableau indicatif des courbes de croissance chez les chiots de grande race

AGE DU CHIOT	SI LE POIDS ADULTE DU CHIOT EST DE									
	25 kg	30 kg	35 kg	40 kg	45 kg	50 kg	55 kg	60 kg	70 kg	80 kg
2 mois	6,5	7	8,5	9	10	11	12	13	14	15
3 mois	10	11	13	14	16	18	19	20	22	24
4 mois	13	15	17	19	21	23	24	26	29	31
5 mois	16	19	21	23	25	28	29	31	35	39
6 mois	18	21	24	27	29	32	34	36	40	44
7 mois	20	23	27	30	33	36	39	41	46	50
8 mois	22	25	29	32	36	39	41	44	50	54
9 mois	23	26	30	34	38	42	45	48	54	59
10 mois	23	28	32	36	40	44	47	50	57	62
12 mois	25	29	34	38	42	47	51	55	63	71
14 mois		30	35	39	44	49	53	58	66	74
16 mois			40	45	50	54	58	67	76	
18 mois						55	59	68	77	
20 mois							60	69	78	
22 mois								70	79	
24 mois									80	



Base de données et système de gestion de bases de données (SGBD)

Les données y sont organisées selon un schéma logique et indexées, pour faciliter la recherche et la récupération d'informations.

Le système de gestion de bases de données permet à un ordinateur de stocker, récupérer, ajouter, supprimer et modifier des données. Il a un effet fédérateur car il :

- Permet de gérer plusieurs bases de données sans être contraint par leurs différents formats
- Permet à plusieurs usagers d'interagir sur ces données, de coupler des données provenant de plusieurs fichiers.

Les requêtes sont la principale façon dont les utilisateurs récupèrent les informations de la base de données.

Pour manier et exposer des données, il faut des outils dédiés (*ici typologie simplifiée*)

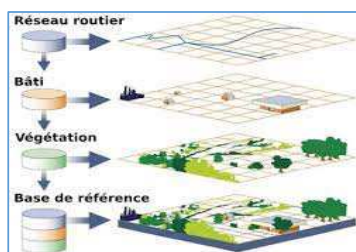
1/2

Data	Comprendre le Big Data à travers les films de cinéma	Débutant
Data	Initiez-vous au Machine Learning	Intermédiaire
Data	Réalisez des calculs distribués sur des données massives	Avancé
Data	Évaluez les performances d'un modèle de machine learning	Intermédiaire
Data	Explorez vos données avec des algorithmes non supervisés	Avancé

Logiciel bureautique

Permet de gérer des données sur le poste de travail, ou éventuellement en partage entre plusieurs utilisateurs.

Exemple : le tableur, qui permet de manipuler des feuilles de calcul (Open Office Calc. , Libre Office Calc, Excel...)



Logiciel spécialisé de traitement de la donnée

Exemple : système d'information géographique, conçu pour recueillir, stocker, traiter, analyser, gérer et présenter tous les types de données spatiales et géographiques (Qgis, ArcGis, GoogleEarth...), généralement utilisable avec les langages Python et R

Exemple : R, qui permet de réaliser des calculs statistiques sur une grande quantité de données et de générer des représentations graphiques

Des centaines pour le pôle ministériel...

Systèmes d'information non projetés en open data (exemples) :

- Chorus (budget)
- RenoIRH (RH)
- Intr@doc (documentation)

Systèmes d'information projetés en open data sur des sites internet (exemples) :

- SI Natura 2000 (biodiversité)
- SI SisFrance (sismicité)
- SI SISPEA (eau et assainissement)

Système d'information (SI)

Ensemble de ressources, à la fois humaines, matérielles et immatérielles, dont le rôle est de collecter, stocker, traiter et distribuer de l'information.

Un SI est un assemblage : des bases de données, des matériels, des logiciels (dont un système de gestion de bases de données), un réseau d'acteurs, des pratiques (pour la qualité de la donnée, la maintenance, la sécurité, etc.). Il peut avoir un système de gouvernance élaboré (rôle défini des acteurs, organisation de la décision, ...).

Des SI peuvent communiquer entre eux, on parle alors d'interopérabilité (avec schéma de données commun, développement d'API entre les SI...). Des SI thématiques peuvent être « intégrateurs » de SI sous-thématiques.

C'est généralement sous cette forme :

- que nous accédons à nos données métier et les complétons, qu'elles aient ou non vocation à être en open data
- qu'est préparée la donnée pour être projetée sur les sites / plateformes internet. D'ailleurs, en général, le système d'information et le site / la plateforme internet portent le même nom

On en compte quelques dizaines pour le pôle ministériel

Il en existe pour l'ensemble de nos thématiques métier

Les données y sont plus ou moins bien « cataloguées » (un catalogue de données exploite les métadonnées et les outils de gestion des données afin de créer un inventaire des actifs de données, permettant aux utilisateurs de les trouver et d'y accéder facilement)

Ils sont plus ou moins bien référencés par les moteurs de recherche

Certains sont des portails qui donnent accès à d'autres portails ou sites et offrent des services pour croiser les données ainsi juxtaposées



Site internet, Plateforme internet, Portail internet, Point d'accès national de données...

NB : tous les sites internet utilisent de la « donnée » numérique. Ici on vise ceux en particulier dont l'objet est d'exposer de la donnée

Ils exposent des données ouvertes des systèmes d'information (et/ou des données partagées c'est à dire protégées au sein de sous-espaces à accès sécurisé avec « administration des droits »)

- en y associant **des services** utiles à l'utilisateur (moteur de recherche, outil cartographique, téléchargement en format ouvert, lecteur de vidéos, API, authentification de l'utilisateur, ...)
- et souvent en faisant « parler » la donnée (« **connaissance** ») avec des datavisualisations, des articles, ...

Ils essaient d'effacer pour l'utilisateur non spécialiste la complexité de la gestion des données. Ils peuvent être plus ou moins interactifs avec l'utilisateur.

C'est généralement sous cette forme que nous prenons connaissance de tous les jeux de données ouverts du pôle ministériel, y compris ceux qui ne sont pas relatifs à notre activité

NB : du point de vue du vocabulaire, généralement,

- le « site » est plutôt mono-thématique
- le « portail » ou la « plateforme » sont plutôt ensembliers car :
 - pluri-thématiques
 - et/ ou points de rassemblement de plusieurs systèmes d'information et leurs bases de données
 - et/ou dédiés à un territoire, un grand projet

Commun de données

Plateforme internet de données à gouvernance collective et largement ouverte au public pour l'enrichissement des données par le plus grand nombre

Exemples : OpenStreetMap, Open Food Facts

Notre pôle ministériel produit un nombre très important de jeux de données (ouvertes ou non)

Fruit de l'histoire *(outre l'histoire générale de l'ouverture des données publiques...)*

- Il existe un droit de toute personne à accéder aux informations relatives à l'environnement : convention d'Aarhus du 25 juin 1998, directive du 28 janvier 2003 concernant l'accès du public à l'information en matière d'environnement, articles L124-1 à L124-8 du code de l'environnement...
- La directive du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne (dite directive Inspire), transposée en 2010 dans le code de l'environnement, oblige à produire et ouvrir des jeux de données géographiques et environnementales
- La production et mise à disposition de certaines données est rendue obligatoire par des textes communautaires ou français :
 - ❖ statistiques, biodiversité, eau, milieux marins, transports, géologie, filières Responsabilité élargie du producteur (de déchets)...
 - ❖ données dites de référence (base adresse nationale, etc.)
 - ❖ Données dites de forte valeur (au 10 avril 2024 - Règlement (UE) du 21 décembre 2022) : géospatiales ; observation de la Terre et environnement ; météorologiques ; mobilité (+ statistiques sur les entreprises et la population)

Fruit de nos missions

- des champs de compétences très étendus et à impact économique et social essentiel justifient de produire de la connaissance
- certains opérateurs ont pour mission de produire et diffuser de la donnée : Météo France, Institut national de l'information géographique et forestière (IGN), Office français de la biodiversité (OFB), notamment
- un important réseau scientifique et technique a pour mission de produire des données de recherche

Fruit de notre organisation

- nous partageons et croisons de la connaissance sur les territoires avec les nombreux ministères, délégataires de service public, collectivités territoriales, associations réalisant des missions de service public entrant dans le champ de compétences ministériel

Et nous aussi nous utilisons beaucoup de données externes :

- fichiers fonciers issus du cadastre de la DGFIP (min Economie)
- cartes marines du service hydrographique et océanographique de la Marine (min Défense)
- données de l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (min Agriculture) et plus généralement de la recherche
- Etc.

Peut-on compter
le nombre de jeux de données du pôle ministériel ?

Question sans réponse

Il suffit de couper en deux un jeu de données pour avoir deux jeux de données... (« jeu de données » n'a pas de définition technique ou juridique)

« *Tout le monde moissonne tout le monde* » (Observatoire OpenDataFrance, rapport mars 2023)
C'est-à-dire que de nombreux sites exposent les catalogues de données issus d'autres sites

Les services peuvent doubler (au minimum) une exposition de données sur un site « métier » et une exposition sur le site national de données publiques data.gouv.fr

Tout ou partie de certains jeux de données ne sont pas en open data : protection de donnée sensible ou personnelle, démarche d'ouverture pas encore menée...

Que signifie « donnée du pôle ministériel » ?

- donnée créée par ses services
- ou donnée relevant de ses (considérables) missions et produite par une myriade de producteurs?

Toujours de nouvelles données : satellitaires, lidar HD sol et sursol, etc.

Peut-on savoir
qui réutilise les jeux de données du pôle ministériel ?

Question sans réponse

Un nombre croissant de données sont récupérées de façon automatique entre systèmes numériques, par API (application programming interface ou « interface de programmation d'application »).

Certaines API permettent toutefois de connaître les coordonnées numériques de la personne / l'institution récupératrice

Le pôle ministériel produit des jeux de données « de référence » largement utilisés par d'autres systèmes d'information : Base Adresse Nationale, Registre parcellaire graphique, Référentiel à grande échelle

On peut éventuellement repérer la première utilisation d'un jeu de données, mais pas les réutilisations de cette première utilisation (utilisations en cascade)

L'utilisation des données ne passe pas forcément par le téléchargement de jeux de données. Il peut passer par des datavisualisations, des applications pour utiliser les données sans les voir, ...

Ex : sur OpenStreetMap on ne sait pas d'où viennent toutes les données qui structurent l'outil

Certaines données sont consultables mais non réutilisables

C'est vrai, il est (encore) compliqué de s'y retrouver !

Comment peut-on améliorer ce paysage foisonnant ? Principales tendances actuelles

Mieux outiller et faire connaître les services Connaissance

Instaurer par un texte un « Point d'accès national » - Ex: transport.data.gouv.fr

Organiser un meilleur repérage des catalogues de données
Faire converger en un seul point les catalogues de données : projet Ecosphères (porté par l'Ecolab avec l'appui de la DINUM)
Ecosphères met en œuvre des logiques de moissonnage permettant la centralisation des métadonnées de près de 28.000 jeux de données appartenant à 137 organisations, actuellement distribuées sur 22 plateformes du pôle ministériel (moissonnage encore incomplet car toutes les plateformes n'ont pas de catalogue)

Organiser un meilleur repérage des systèmes d'information (qui exposent et/ou utilisent les données)
Repérer en un seul point l'ensemble des systèmes d'information du pôle ministériel, leurs utilisateurs majeurs, leurs référents, et les cartographies fonctionnelles (relations) de ces SI :

- projet en cours : GUSI - Gouvernance & Urbanisation du SI (porté par la Dnum). La base (WikiSI) des appli et jeux de données est à disposition des agents sur l'intranet
- Projet plus ancien : le SI des SI par la DREAL Haut de France

Mener vers les données en entrant par les cas d'usage
C'est la 2^{ème} branche du projet Ecosphères : proposer à l'utilisateur des bouquets de jeux de données identifiés selon le besoin professionnel (= cas d'usage) : je mets à jour un PCAET, je conseille un bureau d'études pour un état des lieux de biodiversité, je cherche où implanter un parc d'éoliennes, etc.

Mener vers les données en entrant par la connaissance
C'est l'objectif du portail Notre Environnement (porté par le CGDD au titre du droit d'accès à l'information sur l'environnement), qui propose des articles de présentation d'une thématique et, à la fin, expose des ressources / données « pour aller plus loin »

Réorganiser de façon systémique l'accès et l'usage des données de la transition écologique
C'est un des objectifs du chantier en cours « Numérique et transition écologique » du secrétariat général à la planification écologique (SGPE) (mission de Mme Laura Letourneau) → à suivre

La question qui fâche :

POURQUOI EST-CE QU'ON NE TROUVE PAS TOUTE LA DONNÉE AU MÊME ENDROIT !?!

C'était l'objectif de la plateforme de données nationale data.gouv.fr

MAIS

(et sous réserve d'avis différents bien entendu...)

- Ce serait probablement encore plus difficile d'y repérer sa donnée ... et d'autres communautés renaîtraient alors probablement pour reconcentrer la donnée qui les intéresse
- Ce serait plus compliqué de gérer la donnée partagée non ouverte (droits d'accès)
- La gouvernance (c'est-à-dire qui produit quelle donnée, qui contrôle quoi, qui discute de quoi, qui arbitre quoi, etc.) et le support technique seraient probablement un cauchemar
- Un tel outil serait-il protégé et résilient contre les cyberattaques ?
- Il y a une telle dette technique (une « legacy », un empilement de systèmes techniques de générations différentes) que mettre tous les systèmes d'information et plateformes au même niveau est une gageure
- Il y a d'autres pistes pour s'éviter ces écueils (*v. par exemple diapositive précédente*), sans compter les évolutions des techniques numériques (demain, nous interrogerons peut être directement des outils d'intelligence artificielle pour trouver les bonnes données ?)

Plutôt thématiques

EauFrance

Portail d'accès au système d'information sur l'eau
SI « intégrateur » prescrit par le Code de l'environnement
(DGALN, OFB et autres acteurs)

NatureFrance

Portail d'accès au système d'information sur la biodiversité
SI « intégrateur » prescrit par le Code de l'environnement
(DGALN, OFB et autres acteurs)

Milieu Marin France

Portail d'accès au système d'information milieux marins
SI « intégrateur » prescrit par le Code de l'environnement
(DGALN, OFB, IFREMER et autres acteurs)

transport.data.gouv.fr

Point d'Accès National aux données de transport terrestre, maritime, fluvial, aérien (DGITM)
Prescrit par Règlement UE du 31 mai 2017

Infoterre

Portail géomatique des données géoscientifiques du bureau des recherches géologiques et minières (BRGM)

Géoportail de l'urbanisme

Portail national de l'information réglementaire en urbanisme - Prescrit par Code de l'urbanisme (DGALN-IGN)

Données publiques de Météo France

Base de données nationale des bâtiments (BNBD)
Centre scientifique et technique du bâtiment (CSTB)

Portail open data de l'ADEME

132 jeux de données
Déchets, mesures d'impacts environnementaux, ...

Géolittoral

Portail de la planification maritime et des données mer et littoral (DGALN, CEREMA)

Géorisques

Données sur les facteurs et l'évaluation des risques (DGPR, BRGM et partenaires)

Data Terra

Infrastructure de recherche du domaine système Terre et environnement
(Opérateurs du ministère et d'autres)

Observatoire de la rénovation énergétique, Observatoire de l'artificialisation des sols...

Agence ORE

(Opérateurs de Réseaux d'Énergie)

Acteurs français de la distribution d'électricité et de gaz, guichet unique de la donnée

Plutôt ensembleurs

Portail des données statistiques ministérielles

Changement climatique, Énergie, Environnement, Logement, Transports
(Service des données et des études statistiques - SDES)

Plateformes régionales de données

(Etat-Région ou séparées – Toutes data ou Géodata)
DataSud, Data Grand Est, GéoPal, Open Data Occitanie, GéoNormandie, Géo2France, DatARA...

Géoportail et Géoservices

(Seront progressivement remplacés par la Géoplateforme)
Données et bases géographiques
Institut national de l'information géographique et forestière IGN

Notre Environnement

Portail des informations environnementales des services publics (CGDD)

Observatoire des territoires

Zonages réglementaires, connaissance des territoires (ANCT)

data.gouv.fr

Plateforme ouverte des données publiques françaises (DINUM)
De nombreux jeux de données du pôle ministériel y figurent, soit en dépôt simple soit en doublon
On y trouve en particulier beaucoup de données Energie et Logement-Bâtiment dont la BDNB

NB : le site <https://api.gouv.fr/> référence les API du service public et permet d'accéder aux jeux de données correspondants

Au sein des données ouvertes dans les territoires, on peut distinguer les données d'information géographique ou « géoplateformes » (sur cette carte les plateformes du niveau régional)

Plateformes régionales d'animation de l'information géographique ou « géoplateformes »

En plus de l'action plus ancienne de l'Etat (DREAL et DDT), et depuis la loi du 7 août 2015 portant nouvelle organisation territoriale de la République (loi NOTRe), la Région Collectivité territoriale doit :

coordonner « au moyen d'une plateforme de services numériques (...) l'acquisition et de la mise à jour des données géographiques de référence nécessaires à la description détaillée (du) territoire (...), données dont elle favorise l'accès et la réutilisation »



AU SEIN D'UN ENSEMBLE FOISONNANT

L'association Open Data France recense 250 plateformes territoriales de données

pour l'ouverture des données (toute donnée confondue) dans les territoires (tous niveaux confondus)

Comptage début avril 2023

Plateformes régionales d'animation de l'information géographique (à statut Etat-Région sauf exceptions)



Les pôles « Connaissance » régionaux à vocation nationale

EN CONSTRUCTION

La mission « Connaissance des territoires pour la transition écologique » (ou « Mission Connaissance ») est une initiative lancée en mai 2021 et co-pilotée par le Secrétariat général et le CGDD.

Son objectif est de conforter le rôle de la connaissance dans l'élaboration et la conduite de nos politiques publiques, en tirant parti de nos atouts en termes d'expertise métier et de richesse des jeux de données.

Pour cela, la mission développe une nouvelle organisation et offre de services en coordonnant la création de pôles thématiques « Connaissance » régionaux à vocation nationale.

Chaque pôle est hébergé par un service déconcentré (DREAL ou DDT) et travaille sur une thématique métier.

Ces pôles sont chargés de créer des services et outils de pilotage et d'aide à la décision (tableaux de bord, datavisualisations, cartographies) à partir des jeux de données métier, au service des agents de l'Etat et des collectivités territoriales.

MISSION-CONNAISSANCE ¶



ORGANISATION DES PÔLES ACTUELS



Quelques récentes expositions de données pour le pôle ministériel (en guise d'illustration)

Ouverture de nouveaux jeux de données en 2022

Le registre d'immatriculation des copropriétés

Données techniques et juridiques sur les copropriétés
(Agence nationale de l'habitat)

La base de données nationale des bâtiments (BDNB)

Cartographie et qualifie les bâtiments existants
(résidentiels et tertiaires)
(Centre scientifique et technique du bâtiment)

La carte des loyers

Contient des indicateurs de loyers d'annonces
à l'échelle de la commune
(Ministère de la transition écologique)

(sur la plateforme data.gouv.fr)

Réutilisations de jeux de données en 2022

L'extension navigateur "Encadrement des loyers"

Permet de savoir si une annonce
respecte l'encadrement

Normales de saison des stations météorologiques

Visualisation cartographique

Panorama de l'électricité renouvelable

Usage du vélo dans les mobilités professionnelles

La carte "France Chaleur Urbaine"

Permet de visualiser les tracés de réseaux de chaleur
et les potentiels de raccordement à l'adresse

(sur la plateforme data.gouv.fr)

Trois illustrations de cette présentation générale dans notre webinaire

Plateforme transport.data.gouv.fr

Point d'Accès National aux données de transport qui rassembler les données de toute l'offre de mobilité en France

Plateforme Géorisques

Portail d'information public qui permet d'évaluer son exposition aux risques majeurs sur le territoire français.

Plateforme Géo2France

Mutualise l'information géographique et ouverte en Hauts-de-France

Tout en vous présentant leurs jeux de données et plateforme dédiées, leurs porteurs vont essayer de vous montrer :

- le défi que représente, pour eux aussi, le fait de trouver et récupérer la donnée utile
- les formes que prend concrètement la donnée au sein d'un système d'information
- les métiers et compétences qui se cachent derrière les systèmes d'information
- le dilemme entre :
 - ❖ la facilitation de l'accès à la donnée (s'adresser aux experts en proposant certains formats de données, des API)
 - ❖ et la facilitation de l'accès à l'information (s'adresser aux non-experts en proposant des formats simples de fichiers, etc.)
- l'importance de la gouvernance de la donnée
- quelques techniques et réflexes à avoir pour trouver la bonne donnée.

BONUS

Adresses des sites et références évoqués dans le diaporama

- Plateforme intranet du Wiki SI (les systèmes d'information et jeux de données identifiés dans le pôle ministériel) : <http://wikisi.e2.rie.gouv.fr/wikisi/>
- Portail EauFrance : <https://www.eaufrance.fr/>
- Portail NatureFrance : <https://naturefrance.fr/>
- Portail Milieu Marin France : <https://www.milieumarinfrance.fr/>
- transport.data.gouv.fr : <https://transport.data.gouv.fr/>
- Infoterre : <https://infoterre.brgm.fr/>
- Géoportail de l'urbanisme : <https://www.geoportail-urbanisme.gouv.fr/>
- Données publiques de Météo-France : <https://donneespubliques.meteofrance.fr/>
- Base de données nationale des bâtiments : <https://www.data.gouv.fr/fr/datasets/base-de-donnees-nationale-des-batiments/>
- Portail opendata de l'ADEME : <https://data.ademe.fr/>
- Géolittoral : <https://www.geolittoral.developpement-durable.gouv.fr/>
- Géorisques : <https://www.georisques.gouv.fr/>
- Data terra : <https://www.data-terra.org/>
- Données d'énergie de l'agence ORE : <https://www.agenceore.fr/>
- Observatoire de la rénovation énergétique : <https://www.ecologie.gouv.fr/observatoire-national-renovation-energetique>
- Observatoire de l'artificialisation des sols : <https://artificialisation.developpement-durable.gouv.fr/>
- Portail des données statistiques ministérielles : <https://www.statistiques.developpement-durable.gouv.fr/>
- Géoportail : <https://www.geoportail.gouv.fr/>
- Géoservices : <https://geoservices.ign.fr/>
- Transcription géographique du découpage administratif du territoire français : <https://geoservices.ign.fr/adminexpress>
- Notre Environnement : <https://www.notre-environnement.gouv.fr/>
- Observatoire des territoires : <https://www.observatoire-des-territoires.gouv.fr/>
- data.gouv.fr :
 - ministère : <https://www.data.gouv.fr/fr/organizations/ministere-de-la-transition-ecologique/>
 - données diverses à composante énergie : <https://www.data.gouv.fr/fr/pages/donnees-energie/>
 - données diverses à composante logement et urbanisme : https://www.data.gouv.fr/fr/pages/donnees_logement-urbanisme/
 - données diverses à composante géographique : <https://www.data.gouv.fr/fr/pages/donnees-geographiques/>
- Le SI des SI : <https://www.hauts-de-france.developpement-durable.gouv.fr/?Le-Systeme-d-information-des-systemes-d-information-SIDSI>
- 250 plateformes territoriales de données repérées par Open Data France : <https://airtable.com/shrWxHPi2XjLu9xtM/tblwkJJPsyayeH5IX>
- API publiques : <https://api.gouv.fr/> et Doctrine des API dans les administrations : <https://api.gouv.fr/guides/doctrine-api>

« Données de référence »

« Jeux de données qui présentent le plus fort impact économique et social »

À ce jour, neuf jeux de données, qui couvrent un large champ thématique, ont été identifiés comme des données de référence. Trois sont pilotés par le pôle ministériel de la transition écologique et énergétique et de la cohésion des territoires.

Article L. 321-4
du Code des relations
entre le public
et l'administration

	Pilotage
Base Adresse Nationale (BAN) : la seule base de données d'adresses officiellement reconnue par l'administration	Co-pilotage <ul style="list-style-type: none"> • Direction interministérielle du numérique (DINUM) • Agence nationale de la cohésion des territoires (ANCT) • Institut national de l'information géographique et forestière (IGN)
Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)	Institut national de la statistique et des études économiques (INSEE)
Code Officiel Géographique (COG)	Institut national de la statistique et des études économiques (INSEE)
Plan Cadastral Informatisé	Direction générale des finances publiques (DGFiP)
Registre parcellaire graphique (RPG) : contours des parcelles et îlots culturaux et leur groupe de cultures majoritaire	Co-pilotage <ul style="list-style-type: none"> • Agence de services et de paiement (ministère de l'Agriculture) • Institut national de l'information géographique et forestière (IGN)
Référentiel de l'organisation administrative de l'Etat	Direction de l'information légale et administrative (DILA) (Services du Premier ministre)
Référentiel à grande échelle (RGE) : infrastructure nationale de données géolocalisées qui décrit le territoire français et l'occupation de son sol de façon précise, complète et homogène	Institut national de l'information géographique et forestière (IGN)
Répertoire National des Associations (RNA)	Ministère de l'Intérieur
Répertoire Opérationnel des Métiers et des Emplois (ROME)	Pôle Emploi

Petit glossaire de la donnée

- La **donnée** est l'élément de base de la connaissance. On parle de **jeu de données** ou de **collection de données** pour un recueil de données.
- On parle de **données brutes** pour caractériser des données tout juste acquises (via des appareils de mesure par exemple). On peut **croiser** des données, les **enrichir**, les **agréger**, les **dégrader**, ...
- Une **base de données** est un recueil **structuré** de données.
- On structure une base de données à l'aide d'une **modélisation conceptuelle de données** ou d'un **schéma de données** : pour cela il faut collaborer entre experts de la donnée et experts métiers. Les schémas de données permettent de décrire des modèles de données : quels sont les différents champs, comment sont représentées les données, quelles sont les valeurs possibles etc.
- Un **standard de données** est un schéma de données associé à des règles de gouvernance, l'ensemble étant consensuel et partagé par une communauté ou un écosystème
- Quand on connaît le schéma de données, on peut émettre des **requêtes** sur la base de données pour n'obtenir que les données qui obéissent à nos critères. Les **API** (Application programming interface) sont des applications informatiques qui permettent à d'autres programmes d'obtenir des données selon des requêtes.
- On peut croiser des bases de données si on dispose d'une information identique et discriminante dans chaque base, c'est la **jointure**.
- Les informations qui décrivent les données ou les bases de données s'appellent les **métadonnées** (titre, date d'acquisition, date de mise en ligne, propriétaire, licence, informations sur la qualité, URL de téléchargement, ...)
- Les métadonnées permettent de constituer des **catalogues de données** : les données y sont **exposées**. Ainsi un bureau d'études va consulter les catalogues pour connaître l'état des connaissances sur un domaine qu'il va analyser. Il pourra alors **consommer** les données dont il a besoin. On dit que les données **circulent**, elles sont **partagées**.
- Certaines données sont **ouvertes** (on parle d'**opendata**), d'autres qui portent des informations sensibles ne peuvent être consommées que par des utilisateurs habilités : le partage, ce n'est pas que de l'opendata.
- Les catalogues peuvent communiquer entre eux automatiquement et se mettre à jour des dernières évolutions d'autres catalogue, c'est le **moissonnage**.
- Les données sont **hébergées** au sein **d'infrastructures de données** (portails, systèmes d'information, ...).

On parle de la donnée au sein du pôle ministériel parce que : (rappel du 1^{er} webinaire du 26 janvier 2023)

Notre pôle ministériel, vu sa taille et ses missions, génère et utilise (et doit ouvrir et protéger) **une grande masse de données** très utiles au fonctionnement quotidien de la Nation comme à la réussite de la transition écologique et énergétique (NB : lutter aussi contre les contre-vérités). C'est aussi le cas des collectivités, avec lesquelles notre pôle ministériel, et notamment les services déconcentrés, est en relation pour gérer le territoire. La **planification écologique** du projet France Nation Verte va avoir besoin de s'appuyer sur ces données existantes et à créer

Il existe des dispositifs juridiques spécifiques sur le droit de toute personne d'**accéder aux informations relatives à l'environnement**. Il existe aussi des dispositifs juridiques spécifiques imposant la production et le partage des données des transports, de l'énergie, des valeurs foncières, ...

Certains opérateurs ont **pour mission de produire et diffuser de la donnée** : Météo France, Institut national de l'information géographique et forestière (IGN), Office français de la biodiversité (OFB), notamment

Le pôle ministériel compte en son sein un important **réseau scientifique et technique** qui crée et publie des données de recherche

Le pôle ministériel doit former à la donnée dans ses **écoles** (ENTE, ENTPE, ENPC, ENSG, ENSM, ENSAM, ENAC, INSTN)

Plusieurs grandes politiques du pôle ministériel s'appuient sur des délégataires de service public, dont les données peuvent être qualifiés **d'intérêt général**

La directive du 14 mars 2007 établissant une infrastructure d'information géographique européenne (« **Inspire** ») a eu une très grande importance au sein de nos services pour structurer et ouvrir les jeux de données géographiques et environnementales. Dans les 228 données de références au sens d'Inspire, au moins 220 sont issues du pôle ministériel. Ceci explique que les géomaticiens fassent partie *de facto* du champ des métiers de la donnée (même si une montée en compétences est demandée)

Une grande partie des « **Données de forte valeur** » qui devront être partagées au niveau européen au 10 avril 2024 émaneront aussi de nos politiques publiques ministérielles : géospatiales ; observation de la Terre et environnement ; météorologiques ; mobilité (+ statistiques sur les entreprises et la population)

Plusieurs entreprises appartenant aux secteurs économiques régulés par le pôle ministériel font partie des « opérateurs de service essentiels » et des futurs « secteurs hautement critiques » des directives européennes sur la **cybersécurité** de 2016 et 2022 : transports, énergie, eaux potables et usées

Le pôle ministériel a en charge des enjeux de **régulation** d'usages experts de la donnée : drones, véhicules autonomes, ville « intelligente » (*smart city*), ...

Le pôle ministériel est fortement présent sur des **nouveaux types et usages de données** : satellitaires, lidar 3D, jumeaux numériques, ...

Le saviez-vous ?

- L'agence nationale de la cohésion des territoires (**ANCT**), opérateur de notre pôle ministériel depuis juillet 2022, porte les politiques d'accès des territoires et des personnes au numérique et donc à de multiples usages de la donnée
- **Météo France**, opérateur de notre pôle ministériel, possède deux des plus gros supercalculateurs français pour faire tourner ses modèles : Belenos et Taranis
- L'**IGN** cartographie l'anthropocène par des cartes thématiques sur un nombre limité d'enjeux écologiques majeurs qui rendent compte des changements rapides du territoire et des conséquences sur l'environnement

Principes généraux de qualité de la donnée

Document web facile à lire et librement réutilisable

Ministère de la Transition écologique et de la Cohésion des territoires
Ministère de la Transition énergétique

https://www.ecologie.gouv.fr/sites/default/files/Principes_generaux_qualite_des_donnees_MTE_MTECT_1-1.pdf



Pour aller plus loin sur la donnée et la transformation numérique des administrations

2/3

SAUVEZ LA DATE !

9 juin 2023

**Webinaire Ecolab-CVRH
Culture de la donnée**

« L'intelligence artificielle dans le pôle ministériel »

A distance

29 juin 2023

**Journée de la donnée
du pôle ministériel**

« Connaître et agir au service de la transition écologique »

Auditorium de la tour Séquoia (92-La Défense) et à distance

Pour aller plus loin sur la donnée et la transformation numérique des administrations

3/3



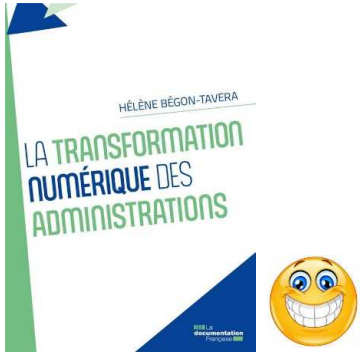
29 juin 2023
Journée de la donnée
du pôle ministériel



Feuille de route de la donnée, des algorithmes et des codes sources
Le Mardi 17 janvier 2023
Pilotée par le Commissariat général au développement durable, la feuille de route propose un cap pour améliorer l'usage de la donnée publique et accélérer la transition écologique sur un horizon de trois à quatre ans. Elle constitue une réponse à la nouvelle politique publique de la donnée inscrite dans la circulaire du Premier ministre du 27 avril 2021 relative à une politique de la donnée, des algorithmes et des codes sources.



Etalab
Politique publique de la donnée
Ouvrir, partager et valoriser les données publiques



Pour une politique publique de la donnée
Rapport
Eric Bothorel - Stéphanie Combes –
Renaud Vedel, décembre 2020

L'Ecolab du CGDD* en quelques mots

Administrateur ministériel des données

pour le Commissaire général au
développement durable

Feuille de route ministérielle de la donnée, des algorithmes et des codes sources

Feuille de route ministérielle de l'intelligence artificielle

EcoSphères : Guichet d'accès à la donnée de la transition écologique

Information géographique : CNIG *, directive Inspire, partenariat avec l'Afigeo

GreenTech Innovation

Labellisation et accompagnement de **start-up et PME** de la GreenTech

Animation du réseau national des **incubateurs** de la GreenTech

Animation du **Living-Lab** de Saint Mandé

Innovation ouverte autour de la donnée pour la transition écologique

Offre d'un bouquet de services sur la donnée : Services Données

Espaces communs de données : Green data for health (GD4H), bio-déchets, ...

Montée des **compétences** sur la donnée

Démonstrateurs d'intelligence artificielle dans les territoires

Structuration de **l'offre de service juridique** sur la donnée

Mission « Connaissance » du pôle ministériel

Etc.



- Commissariat général au développement durable
- Conseil national de l'information géolocalisée