

Mise en place opérationnelle de l'Open Data et inventaire de données

Stéphane Boyera (SBC4D) stephane@sbc4d.com





Objectifs de la session





- 1 Introduction à l'Open Data
- 2 Mise en œuvre d'une dynamique Open Data au sein d'une structure publique
- Rôle et taches d'un Responsable Open Data
- 4 Mise en place d'un inventaire de données



Rappel des Concepts de l'Open Data



Objectifs de l'Open Data





Valoriser les données de l'administration

Elaborer les politiques publiques à partir des données





Intérêt de l'Open Data





- Génération de valeur sociale et financière pour les structures publiques individuellement et par secteur
 - Prises de décision basées sur les données et la réalité
 - Conception et évaluation des politiques publiques
 - Amélioration des services publics
- Génération de valeur pour le gouvernement dans son ensemble
 - Transparence et redevabilité
 - Analyses intersectorielles
 - Efficacité et efficience budgétaire
- Génération de valeurs pour les acteurs non-gouvernementaux
 - Secteur privé (startups, investissements,...)
 - Recherche
 - Media
 - Société civile





Impacts de l'Open Data





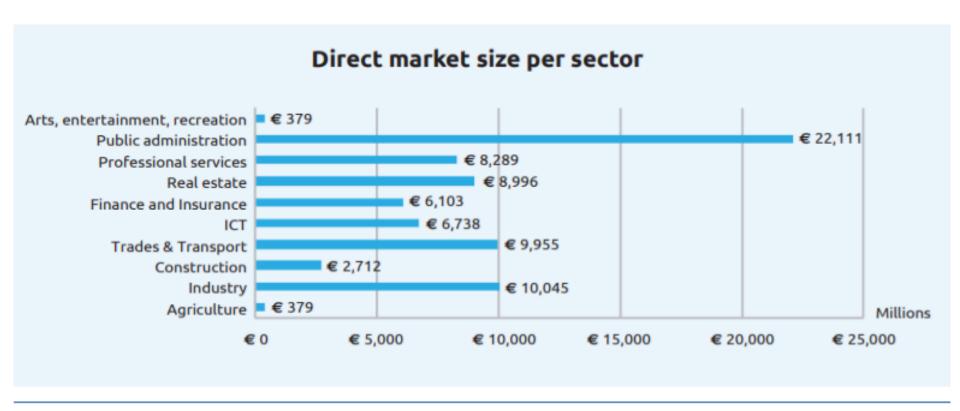


Figure 2 – Direct market size of Open Data per market sector for EU28+ in millions, 2020

Figure 2 - Volume du marché direct des données ouvertes réparties par secteur de l'économie pour l'UE28+ en millions, 2020

L'approche





Les impacts émergent de l'exploitation des données

- L'exploitation de données requiert:
 - L'identification, la documentation, la mise à disposition et le partage d'un gros volume de données de qualité et facile d'utilisation
 - o Des interactions permanentes entre les producteurs et les réutilisateurs de données pour augmenter le volume de données, leur qualité et leur facilité d'utilisation, et pour collaborer sur l'analyse et l'exploitation des données





Toutes les données ne peuvent pas être ouvertes

- Les données liées à la sécurité nationale
- Les données confidentielles
- Les données à caractère personnel sans anonymisation
- Les données couvertes par des limitations légales ou contractuelles
- L'open data ne définit pas les données qui peuvent publiées mais comment les publier pour maximiser leur potentiel de valorisation





Le contexte légal de l'Open Data





- Qu'est-ce que je peux/dois publier?
 - Loi d'accès à l'information: Publication proactive et sur demande
 - Loi de classification des données
- Ou'est-ce que je ne dois pas publier ou qui nécessitent un traitement?
 - Loi d'accès à l'information
 - Loi de protection des données personnelles
 - o Propriété intellectuelle, droits d'auteur et droits voisins
- Quels rémunération je peux demander?
 - Loi d'accès à l'information
- Quelles sont les contraintes de publication?
 - Législation spécifique Open Data
- Qu'est-ce que peuvent/doivent faire les réutilisateurs?
 - Loi d'accès à l'information
 - Législation spécifique Open Data





Données ouvertes: Des données facilement (re)utilisables











1. Données techniquement
ouvertes: disponibles dans un
format standardisé et exploitable
par une machine



2. Données légalement ouvertes:

Ayant une licence d'exploitation autorisant l'utilisation commerciale et non-commerciale et la réutilisation sans restrictions

Données Ouvertes d'un point de vue technique





Une donnée est techniquement ouverte quand:

1

Vous pouvez la chercher et la trouver facilement en ligne

2

Elle est disponible sous un format électronique exploitable ou au travers d'une API

3

Elle est de qualité: à jour, désagrégée,...



xls, json, txt, csv, xml, html, doc, API, odt, ods etc.



PDF, images (JPG, GIF, PNG), d'autres formats propriétaires.



Données Ouvertes d'un point de vue légal (world BANK





Toute personne peut utiliser les données librement Toute personne peut **réutiliser les données** librement Toute personne peut redistribuer les données librement

Pour une utilisation commerciale et non-commerciale On peut demander l'attribution



Copyright © vs





Une licence inclut de nombreux éléments (nonrévocabilité...)





Exemples d'applications

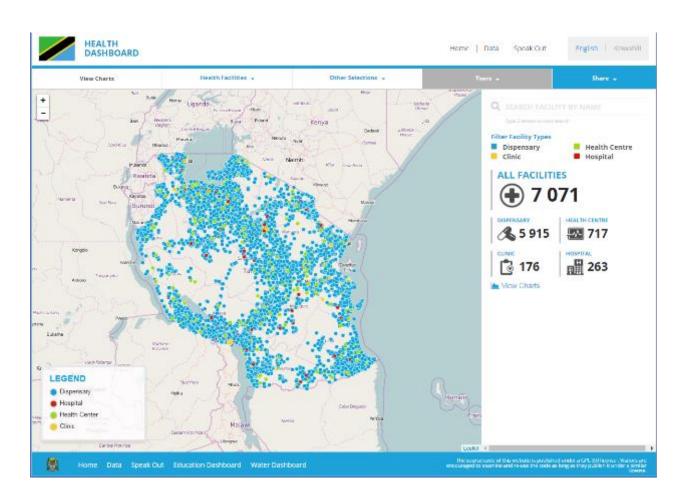


Exemples de service





Centre de Santé en Tanzanie

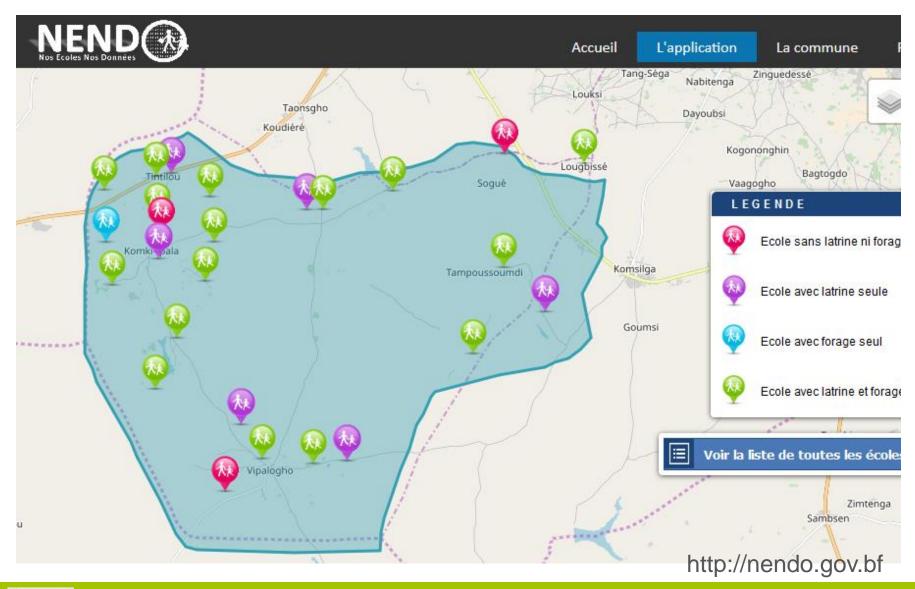




Open Data dans l'éducation





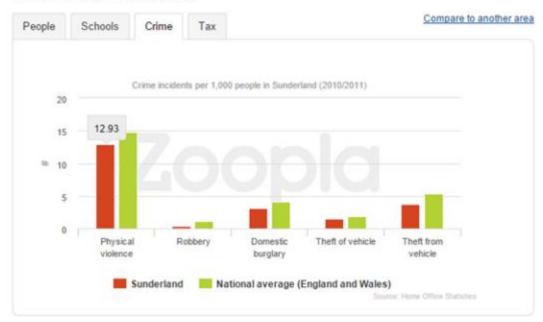




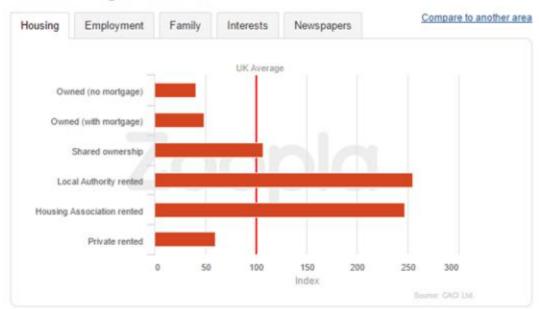
Les futurs résidents peuvent aussi voir:

- Quels sont leurs voisins?
- Le temps moyen pour se rendre au travail
- Le taux de réussite dans les écoles
- La criminalité

Local info for Sunderland



About the neighbours in SR1



Des exemples au Maroc





Comment les marocains organisent-ils leur temps? http://budget-temps.hcp.ma:8081/



Observatoire du marché du travail https://marssad.ma/









Amélioration des services publics





- Le Royaume-Uni a publié la localisation de 300 000 stations de bus au début des années 2010. La communauté a corrigé 18 000 d'entre elles (~6%)
- Priorisation des réparations des routes



https://www.fixmystreet.com

https://teleservices.paris.fr/dansmarue/





Le Responsable Open Data



Les acteurs de l'Open Data







Le ROD et le Responsable de la structure 🛱





- Un leadership fort pour
 - Fixer la vision, les priorités et les objectifs pour l'ensemble des personnels
 - Préparer les mentalités au changement et à l'implémentation de nouvelles approches
- Un Responsable Open Data qui
 - O Crée une dynamique au sein de la structure
 - Permet aux personnels d'acquérir les capacités nécessaires
 - Appuie et collabore avec le CAI et les responsables de la protection des données personnelles
 - Appuie les gestionnaires de données dans la publication
 - Collabore et interagit avec les réutilisateurs internes et externes



Les objectifs





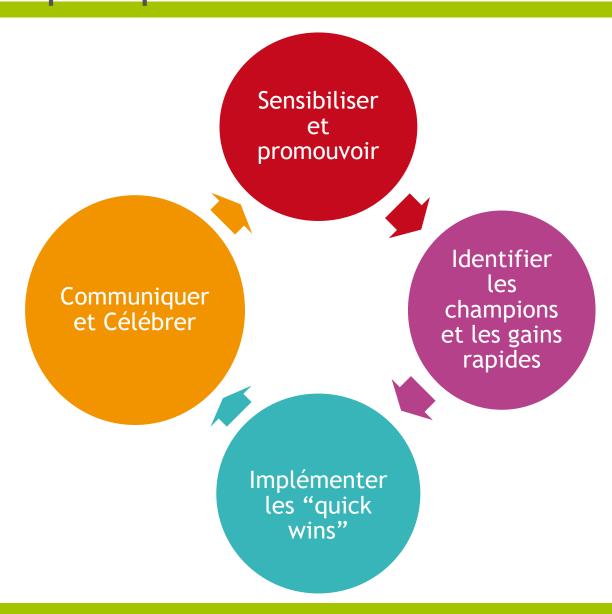
- Une structure publique qui a identifié
 l'ensemble de son patrimoine de données
- Une structure publique qui publie les données qui peuvent être partagées
- Une structure publique qui utilisent ses données et les données des autres structures pour améliorer ses processus de prise de décision, de conception de politiques publiques et d'évaluation
- Une structure publique qui est au contact de ses réutilisateurs externes et qui appuie la réutilisation de ses données pour générer de la valeur économique et sociale



Créer une dynamique au sein d'une structure publique



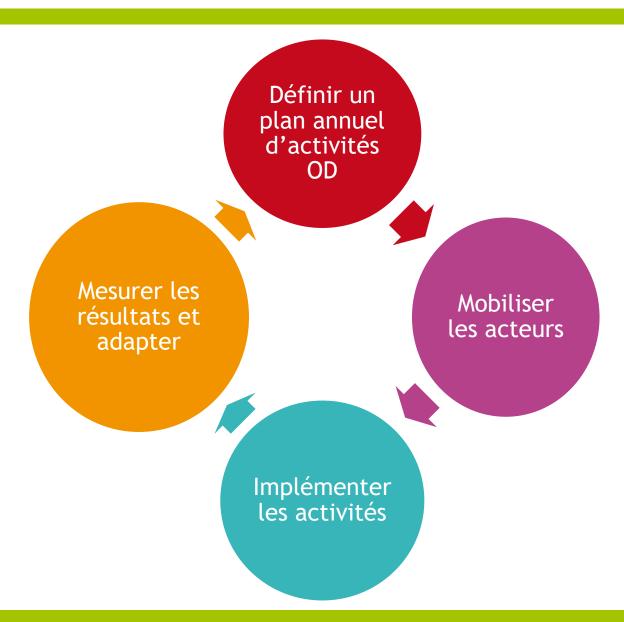




Implémentation de la stratégie OD







Rôle du ROD





- Promouvoir l'Open Data au sein de l'organisme et mettre en œuvre les recommandations du COPIL Open data;
- Définir le plan d'action annuel d'activité Open Data de l'organisme et mettre en œuvre les activités associées;



- Coordonner la publication des données sous format ouvert;
- Vérifier et suivre l'évolution de la qualité des données et en particulier la protection des données personnelles en collaboration avec le Responsable des données personnelles
- Promouvoir les collaborations et les interactions avec les acteurs non-gouvernementaux
- Coordonner le développement et la maintenance de l'inventaire de données de l'organisme;
- Mettre en place et maintenir le plan de suivi et évaluation Open Data de l'organisme;



Le ROD et les gestionnaires de données





Activités	ROD	GDD
Promouvoir l'Open Data	V	
Coordonner les publications	٧	
Préparer et publier les jeux de données		V
Répondre aux questions et commentaires		V
Vérifier la qualité et le suivi des jeux de données	٧	V
Promouvoir la réutilisation et interagir avec les réutilisateurs	V	V
Développement de l'inventaire	V	V





Le ROD et les réutilisateurs





- Promotion de l'Open Data
 - o Blog/Couverture média
 - Présentation/Conférences/Meetup
 - Réunion de sensibilisation
- Mise en place de canaux d'interaction
 - Forum
 - Commentaires sur les jeux de données
 - Requête de jeux
 - Contribution à l'inventaire
 - o Enquêtes
 - O ...
- Développement d'une communauté
 - Référencement à valeur ajoutée: Newsletter, vote,...
- Organisation d'événements d'innovation (hackathon, « demo days »)



Startup



Média





Recherche

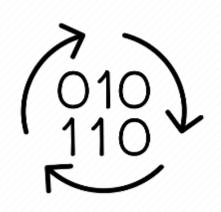


Importance des réutilisateurs





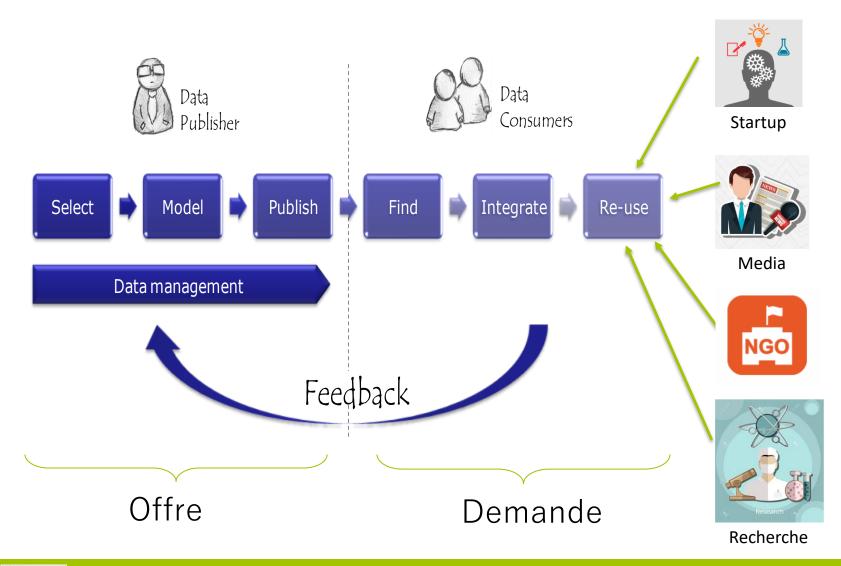
- Améliorer la qualité des données
- Valoriser les données et générer des impacts sociaux et financiers
- Appuyer les structures dans l'analyse des données
- Appuyer les structures dans la mise en place de services publics plus efficaces et plus près des citoyens



Rôle des réutilisateurs



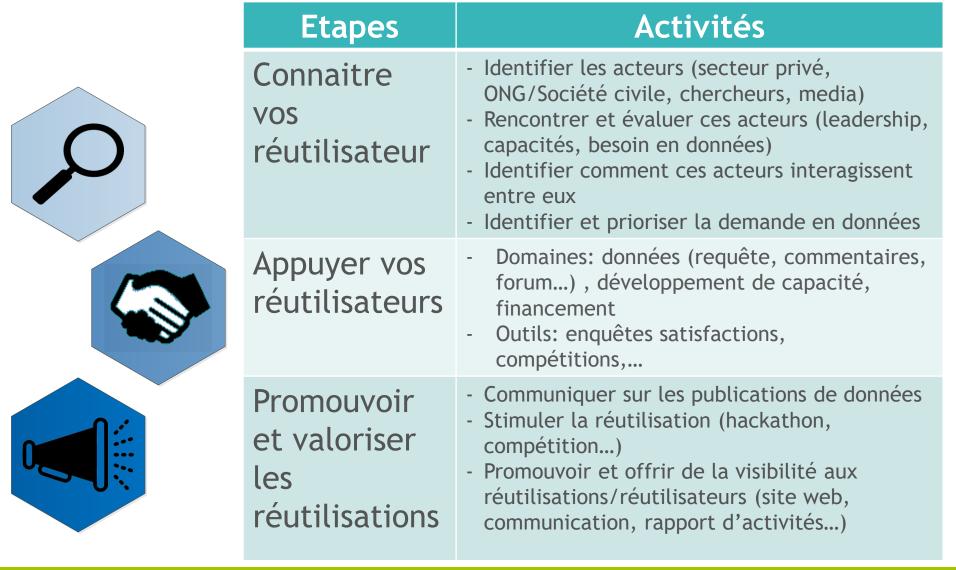




Stratégie d'engagement des réutilisateurs







Facteurs de succès







Produit par OpenDataFrance dans le cadre du programme OpenDataLocale pour l'accompagnement des collectivités à l'open data.





structurer les données le mieux possible pour faciliter leur réutilisation

Documenter et

Construire une vision stratégique avec les décideurs et la communiquer

1

Construire et animer le projet en collaboration avec des partenaires extérieurs et la communauté des réutilisateurs

la durée

Entretenir, enrichir et mettre à jour les données dans

8

Choisir les données en fonction de la stratégie mais aussi des demandes et des opportunités

2

Avancer par étapes et intégrer les retours d'usages pour s'améliorer en continu

5

Favoriser et valoriser la réutilisation des données par tous

(9)

Préférer la qualité à la quantité en optimisant la gestion des données tout au long de leur cycle de vie

3

Publier les données sur des plateformes mutualisées (si possible) et à minima, les référencer



Sensibiliser et former les agents internes à la production et la réutilisation des données

10





Publication de données ouvertes



Rôle du gestionnaire de données





- Préparer et publier les données sur le portail national Open Data
 - Dimension technique
 - Dimension administrative:
 - Niveau d'accès: Inventaire
 - Licences de réutilisation
- Analyser, croiser, visualiser et exploiter les données pour faciliter la prise de décision ou mesurer l'impact des politiques publiques



Gestionnaires de Données

Publication de données





Objectifs:

- Enrichir et mettre à jour l'offre de données ouvertes de l'administration Faciliter la réutilisation des données
- Implémenter la loi d'accès à l'information (Loi 31-13)
 ... dans le respect des autres lois, comme la loi de protection des données personnelles (Loi 09-08) ou la loi sur la cybersécurité (Loi 05-20)



Données vs Jeux de données





- Jeu de données: un ensemble structuré de données qui évolue dans le temps
 - Nouvelle version
 - Correction
 - Date de naissance et date de décès
- Jeu de données = Données + Métadonnées



Dataset

La publication d'un jeu de données





	Etapes	Acteurs	Tâches
Q *	Identifier	ROD - Gestionnaires & Producteurs de données	Rencontrer les agentsCartographier les donnéesDocumenter les données & métadonnées
国	Valider la publication	ROD - Responsables	Evaluer les jeux de donnéesPrioriser l'ouvertureValider l'ouverture
图	Extraire	ROD - DSI - Gestionnaires de données -Prestataires SI	Négocier avec les prestatairesReconstituer schémasDévelopper outils d'extraction
	Editer	ROD - Gestionnaires & Producteurs de données	Expliciter les donnéesAméliorer la qualité & Anonymiser
	Standardiser	ROD - Gestionnaires & Producteurs de données	 Convertir dans un format ouvert Transformer les données selon un standard interopérable
	Publier	ROD - Service Informatique - Gestionnaires & Producteurs de données	- Importer les données sur le portail
3	Maintenir	ROD - Gestionnaires & Producteurs de données	Corriger les erreursMettre à jourRépondre aux questions et commentaires

Préparation des données





Objectif:

- Données structurées
- Données vérifiées
- o Données anonymisées (si nécessaire)
- Données exploitables informatiquement
- Données interopérables
- Données documentées
- Données sous format ouvert







Données désordonnées vs structurées (données lisibles par une machine)





A	Α	В	С	D	Е	F	G	Н	1
1									
2		Monitoring human deve	elo	pment: enla	rgir	ng people's	cho	ices	
3	1 Human	development index							
8									
9		İ							
				Human		Life		Adult litera	icy
10					ent	expectancy	at		•
				index (HD		birth		(% aged 1	
				value		(years)		and above	•
11	HDI ranka			2005		2005		1995-2005	5 ^b
12	HIGH HUN	MAN DEVELOPMENT							
13	1	Iceland		0.968		81.5			d
14	2	Norway		0.968		79.8			d
15	3	Australia		0.962		80.9			d

Country Name	Country Code	Indicator Name	Indicator Code 1	960	1961
Aruba	ABW	Urban population	SP.URB.TOTL	27526	28141
Afghanistan	AFG	Urban population	SP.URB.TOTL	755836	796272
Angola	AGO	Urban population	SP.URB.TOTL	569222	597288
Albania	ALB	Urban population	SP.URB.TOTL	493982	513592

Métadonnées





Des informations sur les données. Ex.

- o Titre Sans ambiguïté
- Description Résumé et description exhaustive
- Langue(s) Langue(s) de publication
- Catégorie Le type de données
- Organisme de publication Qui publie les données?
- Licence d'utilisation Harmoniser les licences!
- o Date de publication, de dernière mise à jour
- Fréquence des mises à jour
- o Format(s) disponible(s)
- Documentation
- > Une donnée sans métadonnées est inexploitable
- > Les métadonnées sont répertoriées dans l'inventaire

Exemple de Métadonnées





Durand	Dupont	12/4	12/2/20
Richard	Laville	6/8	

Nom du marié		Date du mariage	Date de naissance du 1 ^{er} enfant
Durand	Dupont	12/4	12/2/20
Richard	Laville	6/8	

- Registre des mariages de l'état civil de 1923
- Niveau: National/Gouvernorat/ville ...

VS

Registre des mariages de l'église St André de Tunis de 1934

Les standards de données





1. Les formats de fichier

- Importance: Facilite l'exploitation informatique
- Exemples: CSV, XML, Json, GeoJson

2. Les vocabulaires de description des données

- Importance: fournissent de la sémantique aux données
- Exemples: DublinCore, GTFD, OCDS

3. Les normes de représentation

- Importance: Représenter de façon univoque indépendamment de standards ou cultures locales et faciliter l'interopérabilité
- Exemples: norme ISO 8601 pour les dates, norme E.164 pour les numéros de téléphone, norme IEEE 754-2019 pour les nombres décimaux

4. Les identifiants communs (« référentiels »)

- Importance: permet le croisement de jeux de données issus de différentes sources
- un « standard » peut couvrir plusieurs dimensions



Les défis





- Les nombres décimaux: « , » vs « . »
- Les séparateurs de millier
- Les valeurs absentes
- Les dates: YYYY/MM/DD vs JJ/MM/AAAA
- Les numéros de téléphone
- O ...



L'ADD et le CoPIL sont entrain de définir les standards de données

Qualité des Jeux de données





1. Qualité intrinsèque des Jeux de données

- Validité: les métadonnées structurelles doivent intégrer un schéma de validation
- A jour au regard du cycle de collecte
- Données désagrégées
- Données associées à leurs métadonnées administratives, descriptives et structurelles

2. Qualité liée au partage et à l'exploitation des jeux de données

- Niveau d'anonymisation
- Format ouvert
- Données lisibles par une machine
- Facilement identifiable: inventaire, centralisation de la publication
- Données pérennes
- Données Interopérables
- Gratuité / Facilité d'accès
- o Bénéficiant d'un support par le producteur
- L'ADD et le CoPIL sont en train de définir les niveaux de qualité



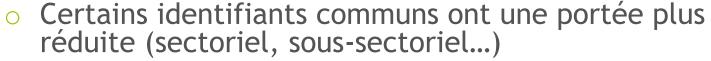


Interopérabilité des données et référentiels de données

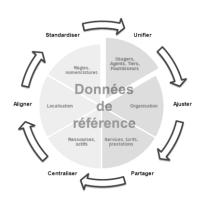




- Le croisement de données de sources diverses requiert l'utilisation d'identifiants communs
- Certains identifiants communs ont une portée nationale
 - Registre des entreprises
 - Base des adresses
 - Base des noms géographiques
 - O ...



- Nomenclature des actes médicaux
- O ...
- L'utilisation de référentiels lors de la publication de données est essentielle
- Le développement d'un référentiel est un processus multi-parties



Anonymisation des données





 Objectif: répondre aux obligations de la Loi 09-08 (protection des données à caractère personnel)

Oritères:

- Est-il toujours possible d'isoler un individu?
- Est-il toujours possible de relier entre eux les enregistrements relatifs à un individu?
- Peut-on déduire des informations concernant un individu?

Processus:

- Les techniques de rédaction
- Les techniques d'agrégation





Les techniques de rédaction





- Suppression d'attributs
- Pseudonymisation
- Généralisation
- Brassage
- > Techniques simples mais relativement peu protectrices.



Les techniques d'agrégation





- Le comptage
- Les totaux
- Les moyennes
- Les distributions : par exemple les quantiles.
- Les agrégations (K-anonymat)
- Les techniques plus avancées: L-diversité,
 T-proximité

Ces techniques dépendent des données, des utilisations, et sont plus difficiles à mettre en œuvre.





Exemple de données personnelles





Nom	Prénom	Sexe	Date de naissance	N° élève	Adresse	résultat	mention
Durand	Paul	M	5/9/2003	1	3 rue de gaulle, 31100 Toulouse	Admis	Bien
Dupont	Marc	M	12/6/2003	2	5 av. Giscard 75000 paris	Non admis	NA
Richard	Josette	F	3/1/2004	3	6 chemin jean 54678 Oeurtil	Admis	Très bien

Durand	Paul	М	5/9/2003	31100	Admis	Bien
Richard	Josette	F	3/1/2004	54678	Admis	Très bien

Α	M	9/2003	31100	Admis	Bien
В	M	6/2003	75000	Non-admis	NA
С	F	1/2004	54768	Admis	Très bien

Importance de l'anonymisation





- Les traitements d'anonymisation doivent avoir lieu AVANT la publication
- o Toutes les données ne nécessitent pas une anonymisation
- Dans de nombreux cas, les techniques simples de rédaction suffisent
- Le traitement des cas où les techniques d'agrégation s'imposent requiert une expertise spécifique qu'il n'est pas forcément nécessaire de développer au sein de chaque structure
- L'inventaire résout les besoins en anonymisation et les techniques à employer
- Le ROD et les gestionnaires de données peuvent demander l'appui de la CNDP et du responsable des données personnelles pour résoudre ces questions et identifier les traitements à appliquer



Qualité vs disponibilité





- Un jeu de données de qualité parfaite n'existe pas
- La mise en place de l'ensemble des critères de qualité retarde la publication et donc l'intérêt et l'impact d'un jeu de données
- Critères essentiels:
 - Validation
 - Métadonnées
 - Anonymisation
 - Format ouvert
 - Données structurées (lisible par une machine)
 - Licences de réutilisation
 - URL pérennes
- Critères importants mais qui peuvent intervenir après la publication
 - Utilisation de référentiels et de nomenclatures
 - Utilisation de standards sectoriels
- Des instructions spécifiques seront rendues publiques sur les critères de qualité après validation du COPIL



Les licences de données





- Les licences: le deuxième pilier de l'Open Data:
 Des données légalement ouvertes
- Les licences fixent les éléments liés à la réutilisation des données:
 - Les droits d'accès
 - Les droits d'utilisation
 - Les droits de partage et de redistribution
 - La non-révocabilité
- Les licences ne concernent que la réutilisation et non le processus de publication:
 - Le niveau d'accès
 - La protection des données personnelles
 - La protection des droits d'auteur et de la propriété intellectuelle
- L'absence de licences ouverte entraine l'absence de réutilisation donc l'absence d'impact

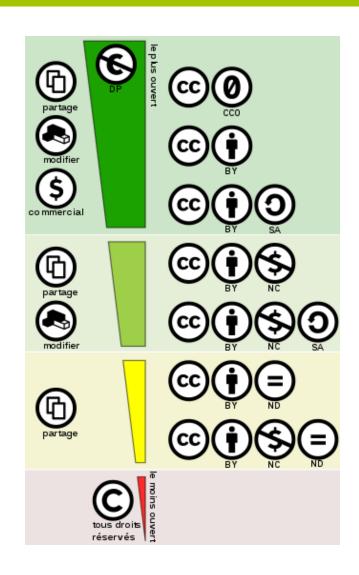


Les différents éléments d'une licence





- Le domaine public
- Les restrictions possibles
 - L'attribution (paternité) : producteur, date
 - Le partage à l'identique (share alike): utilisation de la même licence
 - Utilisation non-commerciale
 - Utilisation non-dérivée
- Les objets des licences:
 - Les données
 - Les bases de données
 - Les codes sources des logiciels
- L'ADD et le CoPIL sont entrain de définir les jeux de licences pour l'Administration



Les éléments qui peuvent influencer la sélection d'une licence





- Flou juridique sur le propriétaire des données
 - Exemple: les données générées lors des marchés publiques
- Exclusivité commerciale
- o La propriété intellectuelle, les droits d'auteurs, ...
- > En dehors de l'attribution, toute restriction supplémentaire limitera l'impact







Inventaire de données



Qu'est-ce qu'un inventaire de données ?





- Une liste exhaustive des données dans la structure
 - Public ou non
 - Qui pourront être publier ou pas dans le futur
 - Digital ou papier
- Un outil pour le ROD
 - Pour la priorisation des publications
 - o Pour l'interaction avec les réutilisateurs
- Un outil pour la structure
 - D'exploitation des données
 - De détection de duplication d'effort
- Un outil pour les autres structures
- Un outil pour les réutilisateurs
- Le terme *inventaire* n'est pas particulièrement adapté. Il s'agit plutôt d'un *annuaire de données*



Pourquoi un inventaire?





- Identifier le patrimoine de données existant;
- Organiser et prioriser la publication des données
- Résoudre l'ensemble des questions administratives liées à la publication de données
- Identifier les redondances dans l'administration et aider à la constitution de référentiels nationaux ouverts.
- L'inventaire de données dépasse le cadre de l'open data
- L'inventaire est un outil de sécurisation des agents et d 'appui à la publication
- La gestion et la maintenance de l'inventaire est une activité continue et pérenne dans le temps

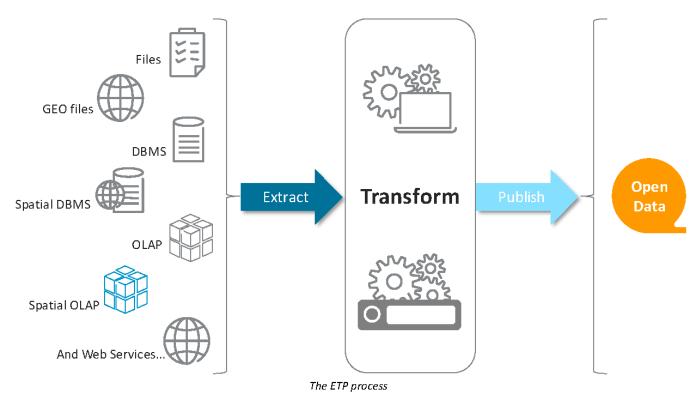
POURQUOI,
POURQUOI?



Processus de publication







https://data.europa.eu/sites/default/files/europe an_data_portal_-_open_data_goldbook.pdf

Processus de l'inventaire







Etapes	Tâches
Identifier les données	Chercher les donnéesCartographier les données
Répertorier les données	- Documenter les données & métadonnées
Classifier les données	Définir les niveaux d'accèsAssocier une licence de réutilisation
Prioriser les données	Identifier la demande interneIdentifier la demande externe

Questions clés et défis





Question	Responsabilité des Structures publiques	Guide de l'inventaire
Qui impliquer à quelle étape ?	\checkmark	
Comment identifier les données ?	\checkmark	
Quelles informations répertorier ?		√
Comment classifier les données ?	\checkmark	
Comment prioriser les données ?	1	
Comment faciliter la préparation des données ?	1	
Où stocker les informations de l'inventaire?		✓
Quelle classification pour l'inventaire?		√



Les acteurs de l'inventaire





Catégorie	Acteurs	Rôle
	Le ROD	Organiser les activités (groupe de travail), Identifier les demandes internes et externes, Classifier
Contributeurs	Le chargé d'accès à l'information	Identifier les demandes externes
	Les gestionnaires de données	Identifier les sources de données, Evaluer la qualité et l'accessibilité Classifier
	Le responsable des données personnelles / de la cybersécurité	Classifier Identifier les traitements
	Le service informatique	Identifier les sources de données Evaluer la qualité et l'accessibilité
Validateur	Le responsable de la structure	Valider la classification
Utilisateur	Le ROD	Répondre aux demandes de données
	Le chargé d'accès à l'information	Répondre aux requêtes
	Le responsable des données personnelles	Gérer les données personnelles
	Les réutilisateurs	Exploiter les données

Bonnes pratiques





- Mobilisation et sensibilisation des personnels impliqués par un haut responsable de la structure
- Nomination officielle des personnes en charge
- Lancement officiel
- Intégration de l'inventaire dans d'autres activités Open Data
 - Développement de cas d'usage internes
 - Développement de cas d'usage externes
 - L'inventaire n'est pas un but mais un outil





Processus de l'inventaire







Etapes	Tâches
Identifier les données	Chercher les donnéesCartographier les données
Répertorier les données	- Documenter les données & métadonnées
Classifier les données	Définir les niveaux d'accèsAssocier une licence de réutilisation
Prioriser les données	Identifier la demande interneIdentifier la demande externe

Quelles données couvrir?







- Favoriser l'exploitation des données par la structure
- Favoriser l'exploitation des données par les autres structures publiques
- Favoriser l'exploitation des données par les acteurs non-gouvernementaux
- Détecter les duplications d'effort
- L'inventaire doit répertorier l'ensemble des données collectées ou produite par la structure (publiable ou non)
- L'inventaire doit répertorier uniquement sur les données primaires collectées ou produites par la structure ellemême (et non les données en provenance d'autres structures)

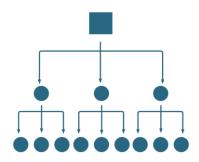


Stratégie de mise en œuvre





- Approche Cyclique
 - Couvrir l'ensemble de la structure par cycle
- Approche stratégique
 - Quelle(s) priorité(s) pour la structure?
 - Approche thématique
 - Approche organisationnelle
- Structure organisationnelle
 - o Inventaire de petite taille : 1 responsable
 - Inventaire de taille moyenne : 1 groupe de travail
 - o Inventaire de grande taille : 1 groupe de travail de responsables de groupe de travail







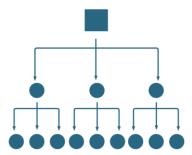
Identification des données





Les sources de données simples à identifier

- Les systèmes d'information et e-services
- Les données déjà publiées (statistiques...) et leurs sources
- Les annuaires (services publics, habilitations...)
- Les données interministérielles (budget, odd, marchés publics...)
- Les données utilisées pour les rapports, les prises de décision, et le suivi & évaluation
- Les données mentionnées dans des accords de partage
- Les demandes d'accès à l'information







Identification des données (2) \(\operation \) WORLD BANK









- L'inventaire est un processus dynamique et itératif
- L'exhaustivité est impossible à atteindre
- On ne fait pas un inventaire mais on met en place un inventaire
 - L'inventaire est en expansion et en évolution constante
 - La maintenance de l'inventaire doit être intégrée aux processus métiers et aux gestions de données



Processus de l'inventaire







Etapes	Tâches
Identifier les données	Chercher les donnéesCartographier les données
Répertorier les données	- Documenter les données & métadonnées
Classifier les données	Définir les niveaux d'accèsAssocier une licence de réutilisation
Prioriser les données	Identifier la demande interneIdentifier la demande externe

Quelles informations répertorier ?





 Le guide méthodologique a arrêté une liste de métadonnées



- Des informations obligatoires
 - Des informations à fournir en français et en arabe (voire en amazigh)
- Des informations optionnelles
- Plus les métadonnées sont complètes plus la publication des données est facilité
- Où stocker ces informations
 - Un outil de capture des informations sous excel (à venir)
 - Constitution d'un jeu de données constituant l'inventaire



Métadonnées





Données à propos des données:

Métadonnées descriptives

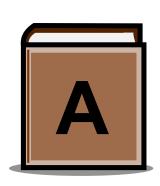
- Permettent de faciliter la recherche et l'identification d'un jeu de données
- o Titre, résumé, auteur, mots-clés
 - Facilite les interactions avec le producteur
- Externes aux données

Métadonnées administratives

- Facilitent la maintenance d'un jeu de données
- Décrivent le contexte de publication:
 - Qui? Quoi? Où? Quand? Comment?
- Décrivent les droits d'accès et de réutilisation
- Externes aux données

Métadonnées structurelles

- Permettent de comprendre les données et de croiser différents jeux de données entre eux
- Décrivent la structure des données (les tables, les colonnes, les indexes)
- Permettent la validation formelle des données
- Internes aux données

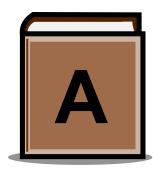




Revue des Métadonnées







Processus de l'inventaire







Etapes	Tâches
Identifier les données	Chercher les donnéesCartographier les données
Répertorier les données	- Documenter les données & métadonnées
Classifier les données	Définir les niveaux d'accèsAssocier une licence de réutilisation
Prioriser les données	Identifier la demande interneIdentifier la demande externe

Comment classifier les données ?





- Classification des données: Définir qui peut accéder à quelles données
- Objectif de la classification:



- A. Qui peuvent être rendues publiques directement
- B. Qui peuvent être rendues publiques après expurgation ou anonymisation
- C. Qui nécessitent des habilitations spécifiques pour pouvoir être consultées (données à accès restreint)
- 2. Identifier la liste des traitements à appliquer pour que les données de catégorie B puissent être rendues publiques
- Les données des catégories A et B (après traitement) entrent dans le cadre de l'Open Data
- Les données des catégories B (sans traitement) et C, peuvent être exploitées par les agents habilités en utilisant les mêmes techniques



Comment classifier les données ?





- Un contexte légal clair
 - Loi n° 31-13 du 22 Février 2018, relative au droit d'accès à l'information



- Loi n° 09-08 du 18 Février 2009, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel
- Loi N° 05-20 du 25 Juillet 2020 (et son décret d'application N° 2-21-406 du 15 juillet 2021) qui définit la classification des actifs informationnels en fonction de l'impact potentiel des incidents de cybersécurité



Approche pratique de la classification des données





o Loi 05-20:

 Chaque entité doit classifier ses données (Article 5) et évaluer si elles font partie d'une des quatre catégories définies (Très secret, secret, confidentiel et diffusion restreinte)

 Les données et actifs informationnels classifiés requiert des habilitations spécifiques pour l'accès et l'utilisation (Article 5)

Loi 31-13:

- Ouverture par défaut (Art. 3): Les citoyens ont accès à toutes les informations sauf exceptions (Art. 7)
- Les exceptions concernent les informations relatives à
 - la défense nationale, la sécurité intérieure ou extérieure,
 - la vie privée des personnes ou les données à caractère personnel
 - les Informations qui portent atteinte aux libertés ou aux droits fondamentaux des personnes



Approche pratique de la classification des données (2)





- Loi 31-13: Les exceptions (cont.)
 - la protection des sources
 - les Informations préjudiciables pour
 - Les relations avec un autre pays ou une organisation internationale
 - A la politique monétaire, économique ou financière de l'Etat
 - Aux droits de propriété industrielle, droits d'auteurs et droits connexes
 - Aux droits et intérêts des victimes, témoins, experts et dénonciateurs dans les affaires de corruption, trafic d'influence et détournement
 - les délibérations du Conseil des ministres et du Conseil du gouvernement
 - o la confidentialités des procédures et enquêtes administratives
 - Le déroulement des procédures juridiques
 - les principes de la concurrence libre, légale et loyale et de l'initiative privée
 - Dans le cadre des exceptions, les informations relevant des exceptions doivent être expurgées et le reste des informations rendues publiques (Art. 8)



Approche pratique de la classification des données (3)





Loi 09-08: « données à caractère personnel » : toute information concernant une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques de son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale;



- Exception:
 - o si la législation prévoit expressément l'enregistrement ou la communication des données à caractère personnel;
 - Si la personne a rendu elle-même les informations publiques
 - D'autres exceptions spécifiques notamment pour la conservation des données à des fins historiques, statistique, ou scientifique
- La présence d'informations nominatives n'empêche pas la publication et nécessite une évaluation spécifique
 - Si les informations doivent être protégées
 - Comment elles doivent être occultées



Diagramme de classification - Loi 05-20





Impact d'un incident de cyber sécurité portant sur la disponibilité, l'intégrité ou la confidentialité des données

Incident à impact très grave: L'incident peut:

- Nuire au maintien des capacités de sécurité et de défense de l'Etat
- Porter préjudice aux intérêts stratégiques de l'Etat
- Porter atteinte à la santé et à la sécurité de la population
- Perturber ou nuire au fonctionnement de l'économie nationale
- Engendrer une incapacité totale ou partielle de plusieurs infrastructures d'importance vitale a assurer leurs fonctions essentielles

Incident à impact grave: L'incident peut engendrer :

- une incapacité totale ou partielle d'une infrastructure d'importance vitale a assurer leurs fonctions essentielles
- Une incapacité totale d'une ou plusieurs entités non considérées comme infrastructures d'importance vitale à assurer leurs fonctions essentielles
- Des pertes financières importantes pour une ou plusieurs entités ou infrastructures d'importance vitale

Incident à **impact modéré**: L'incident peut engendrer :

- Une gêne ou perturbation mineure dans les fonctions d'une infrastructure d'importance vitale
- Une incapacité partielle d'une ou plusieurs entités non considérées comme infrastructures d'importance vitale à assurer leurs fonctions
- Des pertes financières modérées
- Ou toute autre conséquence de nature analogue

Incident à impact limité: L'incident peut engendrer :

- Une gêne ou perturbation dans les fonctions d'une infrastructure non considérée comme infrastructure d'importance vitale
- Des pertes financières limitées
- Ou toute autre conséquence de nature analogue

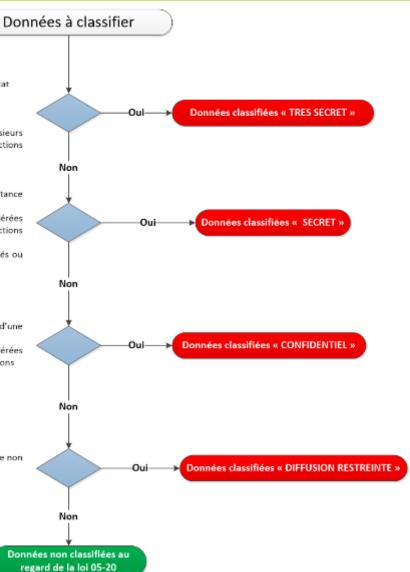
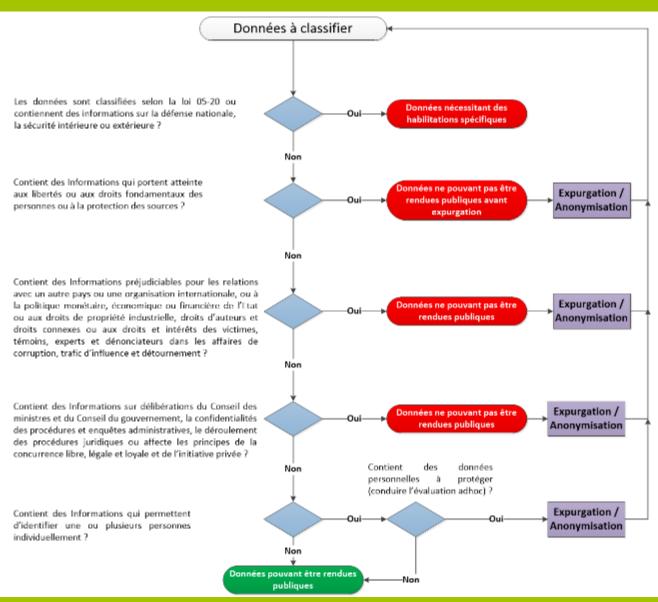




Diagramme de classification - Loi 31-13









Evaluation des données personnelles

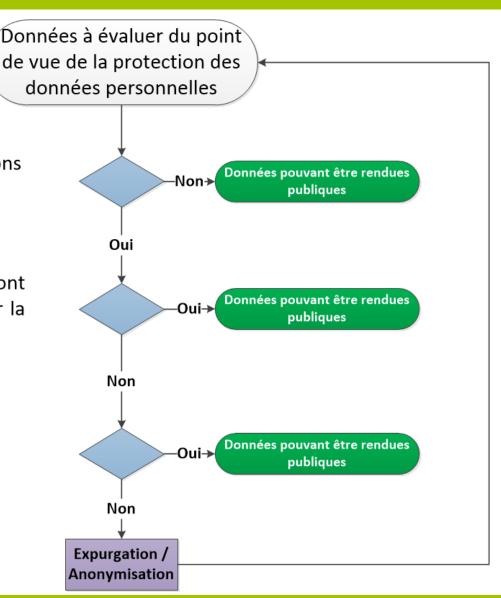




Est-ce que les données Contiennent des Informations qui permettent d'identifier une ou plusieurs personnes individuellement ?

Est-ce que les données ont été rendues publiques par la personne ?

Est-ce que la législation prévoit expressément l'enregistrement ou la communication de ces données ?





Processus de l'inventaire







Etapes	Tâches
Identifier les données	Chercher les donnéesCartographier les données
Répertorier les données	- Documenter les données & métadonnées
Classifier les données	Définir les niveaux d'accèsAssocier une licence de réutilisation
Prioriser les données	Identifier la demande interneIdentifier la demande externe

Comment prioriser les données ?





- Evaluation des jeux de données autour de 3 axes
 - <u>L'utilité et l'impact des données</u>: Est-ce que les données sont utiles :



- En interne
- A d'autres structures
- A des acteurs non-gouvernementaux
- <u>L'accessibilité</u>: Est-ce que les données sont faciles à publier?
 - Elles sont dans un système d'information
 - Elles sont dans un format proche du format de publication
 - Les traitements à appliquer sont faibles et faciles à mettre en œuvre

La qualité

- Les données sont mises à jour de façon systématique
- Les données sont désagrégées
- Les données sont complètes
- Les données sont valides
- Les données utilisent des standards de données et des référentiels



Publier les données de l'inventaire





- Pourquoi publier les données de l'inventaire ?
 - Pour faciliter la réutilisation des données
 - Pour prioriser les publications
- Où publier les données de l'inventaire ?
 - Portail national de données ouvertes
- Quoi publier?
 - Une version publique
 - Une version interne à l'Administration?
 - Une version interne à la structure?



Calendrier d'exécution





Activités		S2	S 3	S4	S5	S6	S7	S8
Constitution d'un groupe de travail								
Identification des cibles								
Sensibilisation & mobilisation des cibles								
Cartographie, Classification & documentation des données								
Priorisation								
Publication de l'inventaire								



Outils & Support





- Un guide méthodologique pour l'inventaire
 - Processus
 - Liste de métadonnées
- Le portail National Open Data







Exemples d'inventaire



Les exemples internationaux







Les Etats-Unis

- Niveau Fédéral
 - L'inventaire couvre toutes les données
 - Vocabulaire dédié <u>DCAT-US Schema v1.1</u>
 - Plateforme par défaut non obligatoire (CKAN)
 - Accès restreint aux informations de l'inventaire
 - Chaque structure est laissé libre dans la mise en œuvre de son inventaire et dans la plateforme



Les exemples internationaux





Les Etats-Unis

- Niveau Communal: San Francisco
 - Une méthodologie précise proposée aux différents départements
 - Une approche hiérarchique en fonction de la taille de la structure
 - L'inventaire couvre toutes les données
 - Un formulaire Excel (pas de vocabulaire formel)
 - Une liste de métadonnées d'inventaire et de métadonnées de publication
 - Pas de plateforme mais <u>un jeu de données</u> <u>public mis à jour annuellement</u>



Les exemples internationaux (2)





Le Royaume-Uni

- Couvre les données publiques uniquement
 - L'inventaire est directement connecté aux requêtes d'accès à l'information
- Utilise le portail national comme plateforme
- Offre la possibilité à tout demandeur de contribuer financièrement à l'ouverture d'un jeu de données particulier (pour l'ensemble de la communauté)



Les exemples internationaux (3)





La Tunisie

- Une plateforme dédiée (PDIMS) (en cours, accès à définir)

- Template Excel
- Un ensemble de métadonnées obligatoires et optionnelles
- Les métadonnées en deux langues (arabe et français)
- Le développement de l'inventaire est coordonnée par le responsable de l'initiative Open Data (UAE) avec une équipe de consultants qui forment les ROD

Ressources





- La documentation sur le portail national : <u>https://data.gov.ma/fr/documentations</u>
 - Les manuels Open Data
 - Les présentations
 - Le plan d'actions national
- La note de l'ADD sur le développement de référentiels
- Le guide méthodologique pour l'inventaire
- Plan d'activités
 - Les Feuilles de route Open Data des Ministères français:
 https://www.data.gouv.fr/en/datasets/feuilles-de-route-ministerielles-sur-la-politique-de-la-donnee-des-algorithmes-et-des-codes-sources/
 - Département du Commerce Etats-Unis:
 https://www.commerce.gov/sites/default/files/2021-08/US-Dept-of-Commerce-Data-Strategy.pdf



THANK Y COU!