

Référentiels des données

14 Décembre 2021 – Séance n°3



UNITED NATIONS

الاسواق
ESCWA

Shared Prosperity **Dignified Life**



Lier les données ouvertes à un référentiel

→ Il est important d'intégrer dans un jeu de données des données pivots relevant d'un référentiel (si elles existent).

Exemple N° 1:

En tant que ROD d'une commune, je souhaite publier un jeu de données contenant la liste des entreprises ayant participé à un appel d'offres pour la réalisation d'un projet spécifique dans ma région.

→ Il est intéressant d'ajouter un champ correspondant à l'identification des sociétés.

Cet identifiant existe, il s'agit de l'ICE « **Identifiant Commun de l'entreprise** ».

Mais, il y a d'autres identifiants légaux : le numéro de registre de commerce, l'identifiant fiscal et le numéro de CNSS !

→ Il faut ajouter l'identifiant, et surtout spécifier le répertoire / référentiel utilisé !

Lier les données ouvertes à un référentiel

Avantages :

- Une meilleure **formalisation**;
- Une meilleure **synthèse**;
- Une meilleure **compréhension**;
- Une meilleure **réutilisation**;
- Une meilleure **interopérabilité**.

<https://guides.etalab.gouv.fr/qualite/lier-les-donnees-a-un-referentiel/#avantages>

Pour quoi un référentiel de données

Pourquoi des référentiels de données ?

- ❑ Les systèmes d'information des acteurs publics sont construits le plus souvent de manière indépendante les uns des autres (sans concertation/ en silo),
 - ❑ Les acteurs publics utilisent souvent leur propre système de codification pour décrire leurs données,
 - ❑ Les structures techniques des données sont souvent différentes,
-
- ➔ Perte de ressources, de temps et d'énergie pour harmoniser et combiner les jeux de données les uns aux autres.
 - ➔ Freine considérablement la réutilisation des jeux de données (surtout l'automatisation de la réutilisation).
 - ➔ Nécessite la mise en place des tables de correspondances (difficile à synchroniser)

Pour quoi un référentiel de données

Principaux enjeux !

- Identification des données de référence :
 - ✓ Identifier de façon unique, pérenne et partagée les objets (c.-a.-d : les données de référence),
 - ✓ Identifier l'unique point de vérité pour toutes les données de référence (l'organisme responsable de la donnée),

- Faciliter le partage et l'interopérabilité :
 - ✓ S'appuyer sur les standards,
 - ✓ Modéliser et formaliser,
 - ✓ Mettre en place un système d'accès aux données de références,

- Faciliter la consolidation et le suivi dans le temps :
 - Mettre en place un système de mise à jour,
 - Prévoir les mécanismes de gestion des erreurs ou de mauvaise qualité,

Exemple de référentiels

- ***Le référentiel national des arrêts***

Identifier de façon unique, pérenne et partagée les arrêts de transport en commun

- ***Le répertoire national des sociétés, associations, etc...***

Lister toutes les informations relatives aux sociétés et associations et les codes d'identification

- ***Les référentiels des codes géographiques***

- Garantir une bonne superposition des couches géographiques,
- Echanger des données géographiques avec des acteurs internes et externes :

Le Cas de La France

Le service public de la donnée (SPD) en France a publié 9 données de référence maintenues de manière pérenne par des administrations et ce sous forme de jeux de données.

Sur la plateforme Open Data « data.gouv.fr » il existe plus de 2000 jeux de données proposant des référentiels de données.

Une grande partie de ces référentiels est proposée par des acteurs non gouvernementaux.

Les données de référence

À ce jour, neuf jeux de données, qui couvrent un large champ thématique ont été identifiés comme des données de référence.

 Base Adresse Nationale (BAN) La Base Adresse Nationale est une base de données qui a pour but de référencer l'intégralité des adresses du territoire 1 33 40 France	 Base Sirene des entreprises et de leurs établissements (SIREN, SIRET) Pour vous abonner à notre lettre d'information Sirene open data actualités, suivez ce lien : 15 7 27 France	 Code Officiel Géographique (COG) Chaque année, l'Insee met à disposition sur son site (insee.fr) le code officiel géographique qui rassemble les codes et 35 13 22 France Commune française
 Plan Cadastral Informatisé (PCI) Plan Cadastral Informatisé PCI Ce jeu de données provient d'un service public certifié Données de référence Le plan cadastral est 3 9 57 France	 Registre parcellaire graphique (RPG) : contours des parcelles et îlots cultureux et leur groupe de cultures majoritaire Le registre parcellaire graphique est une base de données géographiques servant de référence à l'instruction des aides de la 20 2 12 2013-2017 France	 Référentiel de l'organisation administrative de l'Etat Le décret du 14 mars 2017 a institué le Service Public de la Donnée. Celui-ci met à la disposition du public 9 jeux de données 5 4 16 2016-2026 France
 Référentiel à grande échelle (RGE) L'État a confié à l'IGN le développement du référentiel à grande échelle (RGE). Pour ce faire, il fait appel à ses moyens propres 12 2 8 France	 Répertoire National des Associations (RNA) Que contient le RNA ? Le RNA répertorie l'ensemble des associations relevant de la loi du 1er juillet 1901 relative au contrat 19 8 28 1901-2017 France Point	 Répertoire Opérationnel des Métiers et des Emplois (ROME) Dans un contexte marqué par de fortes mutations de l'environnement économique et social, le ROME (Répertoire Opérationnel 5 2 9

Référentiel de données : Méthodologie

Les étapes ci-dessous représentent une démarche à suivre pour la mise en place d'un référentiel de données :

Prérequis : Identifier la donnée de référence ? Une donnée de référence est principalement caractérisée par la stabilité de sa définition et par un fort accès consultatif.

- Mettre en place le mode de gouvernance (pilotage, services métiers à impliquer, etc...);
- Définir le périmètre & la sémantique ;
- Définir le mode d'alimentation, d'actualisation et d'amélioration de la qualité;
- Publier et ouvrir;

Référentiel de données : Méthodologie

I- Mettre en place le mode de gouvernance :

Préciser le mode de pilotage;

Identifier le groupe des experts (INSEA, IMANOR, Equipe de l'open data, etc...);

Identifier la communauté (acteurs externes et internes) producteurs et consommateurs de la donnée;

Assurer la planification et le suivi des travaux ;

Communiquer et diffuser;

→ Une approche itérative de mise en place est fortement recommandée

Référentiel de données : Méthodologie

2- Définir le périmètre & Formaliser la sémantique :

Définir le périmètre : On s'intéresse ici au contenant – la sémantique de la donnée de référence – Définir pour comprendre et se faire comprendre – Monter en abstraction – Le dictionnaire de données (schéma de données) – Préciser l'identifiant à utiliser

Ex : Les adresses des administrations publiques ?

Définir la couverture : Il est question ici du contenu – On s'intéresse aux données , aux instances de l'objet métier (le concept) du référentiel

Définir la portée : pour quel usage le référentiel est il prévu ? On cherche à identifier les acteurs pour lesquels le référentiel peut être utilisé.

Référentiel de données : Méthodologie

2- Définir le périmètre & Formaliser la sémantique :

Pour le choix d'un identifiant → les URLs

<https://{domain}/{type}/{concept}/{reference}>

Ex : <https://reference.data.gov.ma/id/company/000015465321>

{domain} est une combinaison du nom de l'hôte et du secteur concerné.

{type} devrait être l'un d'une petite série de valeurs possibles qui déclarent le type de ressource qui est identifiée. Les exemples typiques incluent: - 'id' ou 'item' pour les objets du monde réel; - 'doc' pour les documents qui décrivent ces objets; - 'def' pour des concepts; - 'set' pour les ensembles de données; - une chaîne de caractères spécifique au contexte.

{concept} pourrait être une collection, le type d'objet du monde réel ou le nom du schéma conceptuel;

{reference} est un élément ou bien concept spécifique.

Référentiel de données : Méthodologie

3- Définir le mode d'alimentation, d'actualisation et d'amélioration de la qualité :

Mettre en place les mécanismes de création et de mise des données (sous forme de services SAO, API Rest, IHM),

Des systèmes d'alimentation collaboratifs pour le cas des référentiels transverses,

Définir les indicateurs de qualité et leur mode de calcul,

L'initialisation et le premier chargement,

La documentation et la formalisation,

Les mécanismes d'ajustement et de collecte des feedbacks

Référentiel de données : Méthodologie

4- Publier & Ouvrir le référentiel de données sous forme d'un jeu de données

Définir les métadonnées descriptives et structurelles,

Joindre la documentation nécessaire,

Associer une licence,

Insérer le jeu de données dans la catégorie de « référentiels de donnée »



Shared Prosperity **Dignified Life**



Discussion