

INTEROPÉRABILITÉ ET PRINCIPES FAIR

Jessie GALASSO-CARBONNEL

19 novembre 2019

Institut de Recherche pour le Développement



INTEROPÉRABILITÉ

Faciles à trouver

Accessibles

> Interopérables

Réutilisables

“L’interopérabilité des données correspond à leur capacité à être intégrées avec d’autres données et à être utilisées et interprétées par des applications et des processus d’analyse, et ce de manière automatique”

Objectif : les données doivent pouvoir être réutilisées et mises en relations avec d’autres données à l’extérieur de l’entrepôt

RECOMMANDATIONS FAIR¹

1. Les (méta)données doivent utiliser un langage de représentation des connaissances formel, accessible, commun et ayant un vaste champ d'application
⇒ Interopérabilité syntaxique
2. Les (méta)données doivent utiliser des vocabulaires qui suivent les principes FAIR
3. Les (méta)données doivent inclure des références vers d'autres (méta)données
⇒ Interopérabilité sémantique

1. <https://www.go-fair.org/fair-principles>

> Interopérabilité sémantique

Associer une signification aux données, les positionner dans un domaine de la connaissance

Inclut le développement de vocabulaires et de schémas pour décrire les données, et de liens entre les données

→ Décrire les données avec des métadonnées

→ Les annoter avec des vocabulaires

Quels schémas de métadonnées, quels vocabulaires utiliser?

> Interopérabilité syntaxique

S'assurer que les formats puissent être lus et interprétés correctement

Concerne les formats techniques qui permettent de stocker et véhiculer les données (leur structure, leur codification...)

Avec quel(s) format(s) représenter les images, les cartes, les vidéos, les données géospatiales, les documents ... ?

Dans quel(s) langage(s) écrire des métadonnées ?

Choisir des :

- Formats de données
- Schémas de métadonnées
- Vocabulaires
- Langages de représentation des connaissances

“Une stratégie de gestion des données doit être esquissée et coordonnée au plus haut niveau possible pour assurer la cohérence des efforts d'interopérabilité”²

Recommandations de normes et standards ?

2. https://ec.europa.eu/isa2/eif_en

Généralistes³

- Encodage de caractères : UTF-8
- Documents : ODF, DocBook, PDF, EPUB3
- Images : ImageGeoTIFF, PNG, JPEG, SVG
- Données Géospatiales : GeoJSON, GeoSpatial-Metadata, GML

Spécifiques à un domaine

- *Wheat Data Interoperability Recommendations* pour les données sur le blé⁴

3. <http://references.modernisation.gouv.fr/interoperabilite>

4. <https://www.rd-alliance.org/>

Généralistes

- DublinCore
- DataCite
- MODS

Spécifiques à un domaine

- Écologie : *Ecological Metadata Language*
- Cristallographie : *Crystallographic Information Framework*
- Sciences de la vie : *Access to Biological Collection Data*

5. rd-alliance.github.io/metadata-directory/standards/

Généralistes

- DBPedia⁶
- Types de ressources⁷
- *Library of Congress*⁸

Spécifiques à un domaine

- Loterre (terminologies scientifiques)⁹
- AGROVOC (alimentation & agriculture)¹⁰

6. <https://wiki.dbpedia.org/services-resources/ontology>

7. http://vocabularies.coar-repositories.org/documentation/resource_types/

8. <https://www.loc.gov/catdir/cpsol/lcco/>

9. <https://www.loterre.fr/>

10. <http://aims.fao.org/fr/agrovoc>

Fournissent une syntaxe pour :

- définir les métadonnées
- définir les schémas de métadonnées
- définir les vocabulaires

Langages standards

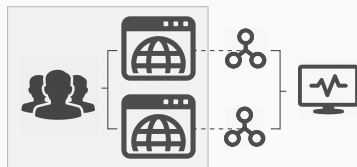
- **XML**¹¹ → faciliter l'échange de contenu complexe entre systèmes hétérogènes
- **RDF**¹² → décrire formellement les ressources Web pour faciliter leur traitement automatique

11. <https://www.w3.org/XML/>

12. <https://www.w3.org/RDF/>

Étendre le web actuel pour rendre son contenu interprétable par des machines

Rajout d'informations structurées
traitables de manière automatique



Définition de standards encourageant l'utilisation de :

- protocoles d'échange sur le web
- **formats de données s'appuyant sur le modèle RDF**

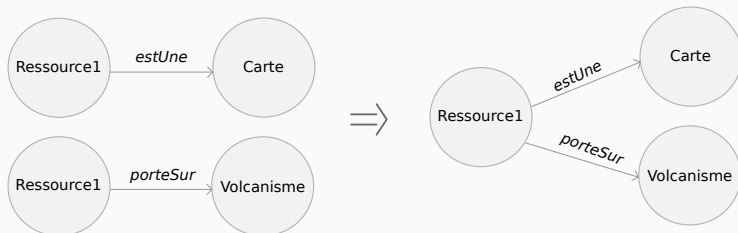
RESOURCE DESCRIPTION FRAMEWORK

Cadre pour décrire les données et les métadonnées sur le web

Informations structurées sous la forme de **triplets**

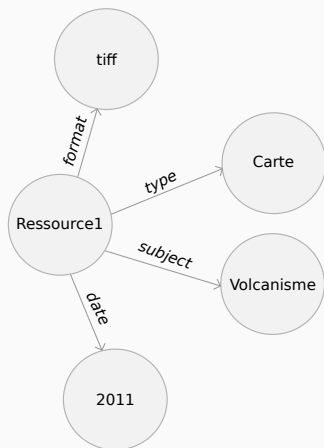
1 triplet = sujet - prédicat - objet

= 2 ressources reliées par une relation



RDF POUR DÉCRIRE LES DONNÉES

Utiliser les propriétés définies dans les schémas de métadonnées



Exemple : DublinCore

- creator
- **date**
- **format**
- isReferencedBy
- license
- **subject**
- title
- **type**
- ...

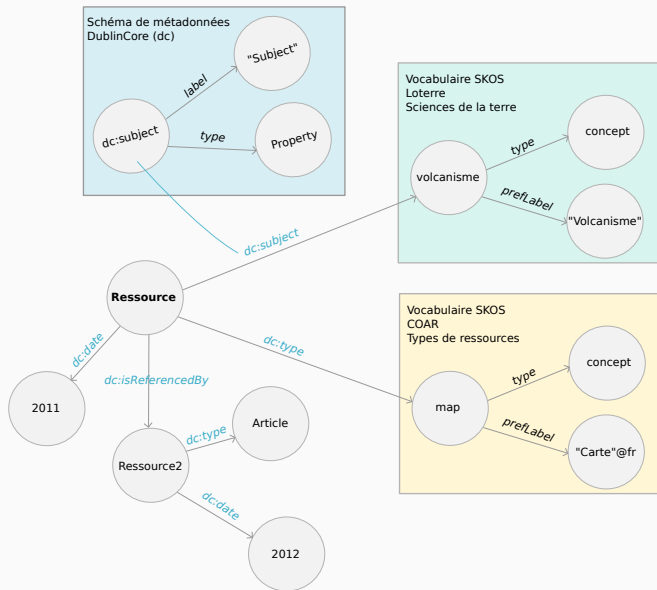
- . Définir des schémas de métadonnées

 - lister les propriétés

- . Définir des vocabulaires

 - **SKOS** : modèle basé sur RDF pour représenter des vocabulaires

 - concepts, termes, synonymes, généralisation, spécialisation, association



Interopérabilité sémantique

- Réutiliser des vocabulaires FAIR existants

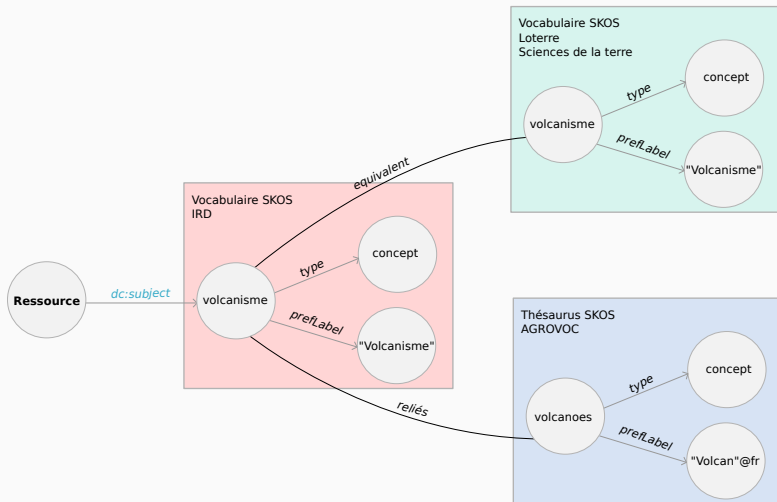
Cas de l'IRD :

- Institut pluridisciplinaire
 - vocabulaires externes trop spécifiques ou trop généralistes
- Répartition géographique spécifique
 - pays des Suds
- Vocabulaires internes anciens

Travail conséquent pour utiliser ses propres vocabulaires

- Identifier les vocabulaires utilisés
 - État des lieux des systèmes d'information
- Les formaliser
- Les préserver
- Les maintenir et les faire évoluer
- Les partager / exposer
- Les aligner

Quels sont les outils qui peuvent nous aider dans ces activités ?



14 trouvés, 6 conservés après avoir écartés ceux n'étant pas open source, pas maintenus ou avec peu de documentation

- . **VocBench3** : complet pour la gestion de vocabulaires
- . **SKOMOS** : complet pour l'exposition de vocabulaires
- . **OpenTheso, Ginco, IQVoc et ThManager** : corrects pour les deux activités

VocBench3 + SKOMOS

Exemples d'utilisation

- Lier données, publications, chercheurs, logiciels
→ VIVO/VITRO¹³ (recherche, visualisation)
- Moteur de recherche
→ Données de sources différentes
- *Discovery tools*
→ Croiser des données, sortir des connaissances
- *Result dashboards*
→ Vue d'ensemble de l'activité scientifique
→ MARLO¹⁴

13. <https://duraspace.org/vivo/>

14. <https://marlo.cgiar.org/ccafs/dashboard.do>