



**HAL**  
open science

# INSPIRE -FAIR PRINCIPLES - OPEN DATA - OPEN SCIENCE

Hélène Bressan, Sylvain Grellet

► **To cite this version:**

Hélène Bressan, Sylvain Grellet. INSPIRE -FAIR PRINCIPLES - OPEN DATA - OPEN SCIENCE. Forum OGC France : Journée Interopérabilité et Innovation (J2I), OGC France; Commission nationale de l'information géographique et spatiale de l'AFNOR, Jan 2024, Champs-sur-Marne, France. hal-04455165

**HAL Id: hal-04455165**

**<https://brgm.hal.science/hal-04455165v1>**

Submitted on 13 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*



Géosciences pour une terre durable

**brgm**

# INSPIRE – FAIR PRINCIPLES

## OPEN DATA – OPEN SCIENCE



Sylvain Grellet – BRGM  
2024-01-23 – Forum OGC France



# Règles INSPIRE

S'appliquent aux données géographiques  
mais aussi données d'observation



## Discovery : Services de recherche

- Catalogues en ligne répertoriant les métadonnées et équipés d'un moteur de recherche, permettant d'identifier des données en fonction de divers critères (thème, mots-clefs, situation géographique, producteur...)



## View : Services de consultation

- Services permettant de visualiser en ligne les données, à l'écran



## Download : Services de téléchargement

- Service permettant de récupérer les données elles-mêmes, pour pouvoir les traiter et les réutiliser



Pour que les données puissent être **publiées et échangées, comprises et réutilisées par les humains ET les machines**, il est nécessaire qu'elles respectent des **règles d'interopérabilité et utilisant des standards**, notamment :

- **Sémantique** : définir, grâce à un modèle de données, le sens, le contenu et la structuration des données.
- **Technique** : utiliser les services web / API reconnus par INSPIRE (services en réseau) et pour lesquels des recommandations ('guidelines') ont été définies avec la communauté
- **Géographique** : présenter les données dans le système géodésique RGF93 .



INSPIRE : s'inscrit dans une dynamique de réutilisation des normes ISO 191xxx et des spécifications OGC

Géosciences pour une Terre durable

**brgm**

\* <https://inspire.ec.europa.eu/data-specifications/2892>

# Principes FAIR

S'appliquent aux données  
et aux métadonnées

## F

**Findable : facile à trouver pour les humains ET les machines**

- F1 : Données et métadonnées ont un identifiant unique et pérenne
- F2 : Données décrites par des métadonnées riches
- F3 : Données et métadonnées enregistrées ou indexées dans un dispositif permettant de les rechercher
- F4 : Métadonnées contiennent l'identifiant unique et pérenne des données



## A

**Accessible : une fois les données trouvées, l'utilisateur doit savoir comment y accéder**

- A1 : Données et métadonnées accessibles par leur identifiants via un protocole standard
  - A1.1 : Le protocole utilisé est ouvert, libre et peut-être implémenté de manière universelle
  - A1.2 : Le protocole peut gérer des procédures d'authentification et d'autorisation si nécessaire
- A2 : Métadonnées accessibles même si les données ne sont plus disponibles



## I

**Interopérable : les données sont interopérables, intégration avec d'autres données, interopérabilité avec des applications**

- I1 : Données et métadonnées utilisant un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
- I2 : Données et métadonnées utilisant des vocabulaires suivant eux-mêmes les principes FAIR
- I3 : Données et métadonnées contenant des liens vers d'autres (méta)données



## R

**Reusable : les données doivent être réutilisables pour de futures recherches**

- R1 : Données et métadonnées richement décrites, par une pluralité d'attributs précis
  - R1.1 : Fournies avec une licence d'utilisation claire
  - R1.2 : Associées à leur provenance
  - R1.3 : Suivent les standards des communautés



<https://force11.org/info/the-fair-data-principles/> RDÂ revolves around this  
<https://www.rd-alliance.org/> and now other organizations (OGC included FAIR in its communications)



# Findable



## F

### Findable : facile à trouver pour les humains ET les machines

- F1 : Données et métadonnées ont un identifiant unique et pérenne
- F2 : Données décrites par des métadonnées riches
- F3 : Données et métadonnées enregistrées ou indexées dans un dispositif permettant de les rechercher
- F4 : Métadonnées contiennent l'identifiant unique et pérenne des données

- F1: Données et métadonnées ont un identifiant unique et perenne (PID)
  - PID = mécanisme permettant d'assigner un identifiant externe stable unique aux ressources du web (Ex: URI, DOI, ...). Permet d'attribuer un identifiant, une reference perenne à ces ressources.
- F2 : Données et métadonnées décrites par des métadonnées riches
  - Pour les réaliser, il est recommandé d'utiliser des normes sémantiques de métadonnées spécifiques (ISO 19115, DCAT) et des formats ouverts (XML, JSON, sérialisations du web sémantique)
- F3 : Données et métadonnées enregistrées ou indexées dans un dispositif permettant de les rechercher
  - Mise en œuvre : utilisation de catalogues et entrepôts de données. Les métadonnées consultables sont mises à disposition afin de permettre leur recherche et celle des jeux de données et faciliter par la suite leur réutilisation.
- F4: Métadonnées contiennent l'identifiant unique et pérenne des données
  - Critère permettant de « passer facilement » des métadonnées aux données et à leurs ressources en ligne.

# Accessible

## A

**Accessible : une fois les données trouvées, l'utilisateur doit savoir comment y accéder**

- A1 : Données et métadonnées accessibles par leur identifiants via un protocole standard
  - A1.1 : Le protocole utilisé est ouvert, libre et peut-être implémenté de manière universelle
  - A1.2 : Le protocole peut gérer des procédures d'authentification et d'autorisation si nécessaire
- A2 : Métadonnées accessibles même si les données ne sont plus disponibles

### ● A1: Protocole de communication standard

- Protocole = procédures utilisées par des machines pour communiquer entre elles
- Les protocoles HTTP et FTP sont des exemples

### ● A1.1: Protocole libre et ouvert

- Le protocole utilisé doit être librement utilisable et interopérable (ex: HTTP, FTP, ...). Les protocoles propriétaires et fermés sont à éviter.

### ● A1.2: Prise en charge de systèmes d'authentification si nécessaire

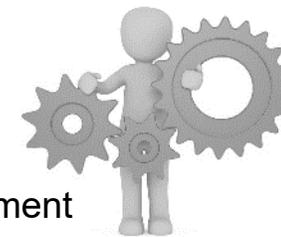
- Accessible n'implique pas toujours un accès libre et ouvert à tout moment. Les données peuvent avoir un accès restreint pour des raisons juridiques (ex : données sensibles, protection des espèces menacées, ...). Ainsi les protocoles de communication permettront d'appliquer des restrictions (ex : HTTPS, FTPS, ...)

### ● A2: Les métadonnées restent accessibles, même quand les données ne le sont plus

- Les métadonnées permettent alors de connaître l'existence de ces données et de contacter les gestionnaires et fournisseurs de ces données. Elles peuvent permettre également de retrouver les articles, publications scientifiques associés.



# Interoperable



**Interoperable : les données sont interopérables, intégration avec d'autres données, interopérabilité avec des applications**

- I1 : Données et métadonnées utilisant un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
  - Représentation des connaissances = modèles de données (UML), ontologies (vocabularies)
  - En 2024, il existe de nombreuses normes et bonnes pratiques internationales à réutiliser. Elles doivent l'être.
  - Ces utilisations permettent de fournir des ressources compréhensibles pour les humains ET les machines. Ensuite, les (méta)données peuvent être échangées de manière transparente (sans aucune intervention humaine pour « câbler/coder » les éléments pour qu'ils fonctionnent)
- I2 : Vocabulaires suivants les principes FAIR
  - = vocabulaires documentés et accessibles par un PID, possédant des métadonnées et étant lisibles par les machines
- I3 : Liens et références vers d'autres données et métadonnées
  - Les principes FAIR sont fortement basés sur les normes et les bonnes pratiques du Web des données
  - Étape par étape, cela crée un réseau mondial d'informations scientifiques reliant les données sources (brutes) à d'autres basées sur elles.
  - La recherche de (méta)données est alors améliorée et permet de découvrir d'autres données

# Reusable



## R

**Reusable : les données doivent être réutilisables pour de futures recherches**

- R1 : Données et métadonnées richement décrites, par une pluralité d'attributs précis
  - R1.1 : Fournies avec une licence d'utilisation claire
  - R1.2 : Associées à leur provenance
  - R1.3 : Suivent les standards des communautés

- R1 : Les méta)données ont une pluralité d'attributs précis et pertinents
  - Plus les (méta)données sont décrites avec de nombreux attributs associés plus la compréhension et la réutilisation peuvent être maximisées.
- R1.1 : Licence d'utilisation des données et métadonnées claires
  - Permet de définir explicitement les droits d'utilisation pour les ré-utilisateurs
  - Le choix de la licence dépend fortement de la nature juridique de chaque donnée (leur contexte d'acquisition, etc...)
  - Il existe des licences gratuites adaptées à l'Open Data [Creative Commons](#), [Open Licence](#), [Open Database License](#)
- R1.2 : Informations de provenance
  - Les informations de provenance, de filiation ('lineage') sont cruciales pour prouver la fiabilité des (méta)données, pour évaluer la qualité des données et permettre de reproduire leur création (ex : expérience scientifique). Celles-ci doivent être fournies lors de la déclaration des données dans une ressource consultable (comme d'autres informations : auteur, institution, dates, ...)
- R1.3 : Répondre aux normes communautaires pertinentes pour le domaine
  - Les standards communautaires permettent d'avoir des données suivant un modèle commun pour faciliter l'intégration des données au sein d'une communauté
  - Si aucune norme communautaire n'existe, la meilleure façon de partager les (méta)données doit être mise en place (Re : principes FAIR)



# Règles et principes d'un vocabulaire FAIR

Ten simple rules for making a vocabulary FAIR (<https://doi.org/10.1371/journal.pcbi.1009041>)

**Rule 1. Déterminer les modalités de gouvernance et le garant du vocabulaire**

**Rule 2. Vérifiez que la licence du vocabulaire autorise la réutilisation et préférer une licence ouverte**

**Rule 3. Vérifier l'exhaustivité et la cohérence des termes et des définitions dans le vocabulaire**

**Rule 4. Établir la traçabilité et l'historique du contenu du vocabulaire**

**Rule 5. Attribuer un identifiant unique et pérenne au vocabulaire et à chaque terme/Concept du vocabulaire**

**Rule 6. Créer des représentations lisibles par machine des termes du vocabulaire**

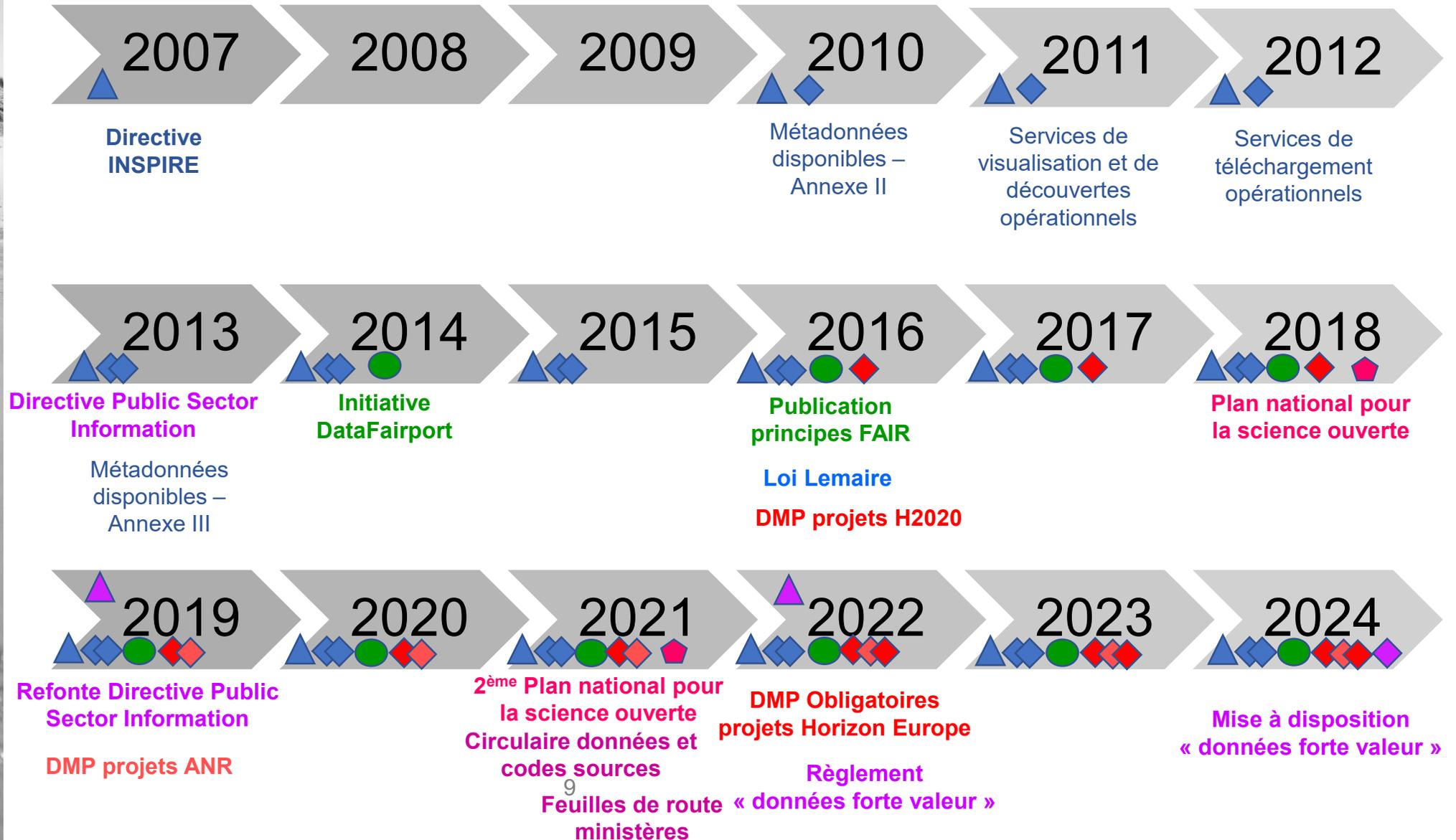
**Rule 7. Ajouter des métadonnées au vocabulaire**

**Rule 8. Enregistrer et déclarer le vocabulaire dans un "Semantic Repository"**

**Rule 9. Rendre le vocabulaire accessible aux humains et aux machines**

**Rule 10. Mettre en œuvre un processus de publication des révisions du vocabulaire**

# Historique et contexte réglementaire





# Conclusion

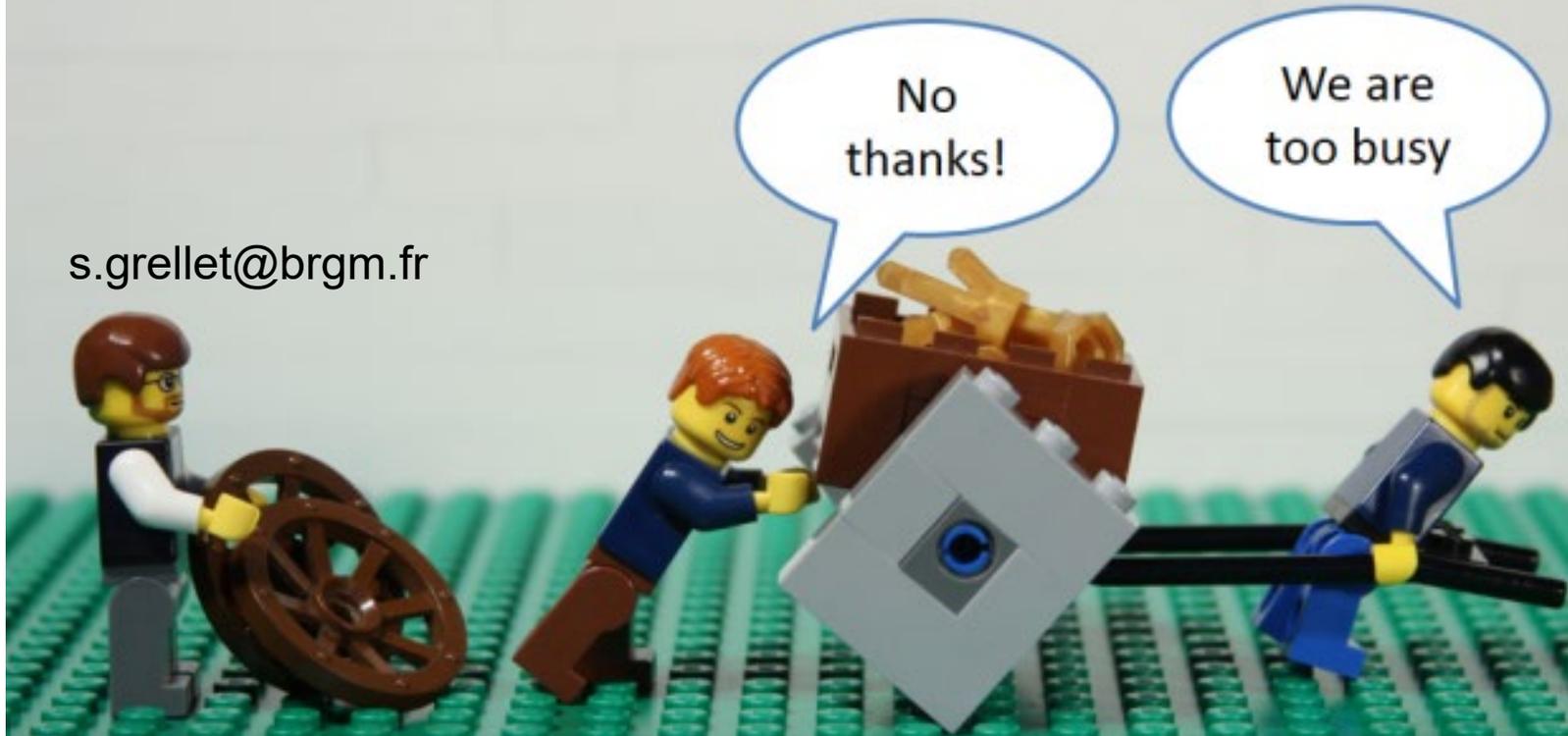
- Un continuum vers un accès plus ouvert et interopérable aux données
- La recherche et le domaine public doivent se conformer à ces dynamiques
- Ces dynamiques s'appuient fortement (sinon uniquement) sur les normes d'interopérabilité convenues au niveau international et les bonnes pratiques établies par des communautés telles que W3C, OGC, RDA, INSPIRE.
- Les exceptions prévues par la loi (au moins en France) ont permis à certaines organisations de modifier leur modèle économique qui reposait auparavant sur la vente de ces données.

**Merci**



Are you too busy to improve?

s.grellet@brgm.fr



Håkan Forss @hakanforss <http://hakanforss.wordpress.com>

This illustration is inspired by and in part derived from the work by Scott Simmerman, "The Square Wheels Guy" <http://www.performancemanagementcompany.com/>



# Merci à

Hélène Bressan

[h.bressan@brgm.fr](mailto:h.bressan@brgm.fr)

