



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



Géosciences pour une Terre durable

brgm

INSPIRE – FAIR PRINCIPLES

OPEN DATA – OPEN SCIENCE



Sylvain Grellet – BRGM
ENSG – mastère GDM – 2023-01-20



INSPIRE Implementing Rules*

Initially for geographic data but also
observation data
- from public organizations

Discovery service

- Services making it possible to search for spatial data sets and services on the basis of the content of the corresponding metadata and to display the content of the metadata



View service

- Service that makes it possible, as a minimum, to display, navigate, zoom in and out, pan or overlay viewable spatial data sets and to display legend information and any relevant content of metadata.



Download service

- Service that enables copies of spatial data sets, or parts of such sets, to be downloaded and, where practicable, accessed directly.



For data to be **published, exchanged, understood and reused by both human AND machines** they need to **respect interoperability rules using standards**

- **Semantics:** defined in INSPIRE themes data specifications (UML) and clarifying the meaning and structure of data (with a legal weight)
- **Technical:** using web services / API endorsed by INSPIRE technical guidelines (network services, many OGC services) defined jointly by the community
- **Geographic :** expose data in coherent spatial reference systems (ex : RGF in France)



=> INSPIRE dynamics reuse a lot ISO191xx approach and OGC specifications

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

FAIR* principles

Applies to data and metadata
- from research organizations

F

Findable : easy to find for human AND machines

- F1: (meta)data are assigned a globally unique and eternally persistent identifier
- F2 : data are described with rich metadata.
- F3 : (meta)data are registered or indexed in a searchable resource
- F4 : metadata specify the data identifier



A

Accessible : once found the user (humand AND machine) needs to know how to access data

- A1 : (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 : the protocol is open, free, and universally implementable
 - A1.2 : the protocol allows for an authentication and authorization procedure, where necessary
- A2 : metadata are accessible, even when the data are no longer available



I

Interoperable : to allow data use regardless of the IT environment used

- I1 : (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2 : (meta)data use vocabularies that follow FAIR principles
- I3 : (meta)data include qualified references to other (meta)data



R

Reusable : data need to be reusable for future research

- R1 : (meta)data have a plurality of accurate and relevant attributes
 - R1.1 : (meta)data are released with a clear and accessible data usage license
 - R1.2 : (meta)data are associated with their provenance
 - R1.3 : (meta)data meet domain-relevant community standards



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





F

Findable: easy to find for human AND machines

- F1: (meta)data are assigned a globally unique and eternally persistent identifier
- F2 : data are described with rich metadata.
- F3 : (meta)data are registered or indexed in a searchable resource
- F4 : metadata specify the data identifier

- F1: data and metadata have a PID (persistent identifier)
 - PID = mechanism allowing to assign an (externally) stable identifier to web resources (ex : URI, DOI, ...). Allows to provide a persistent reference to those resource
- F2: metadata = data describing data
 - To do them it's recommended to use specific metadata semantics standards (ISO 19115, DCAT) and open formats (XML, JSON, semantic web serializations)
- F3: searchable resource -> (meta) data catalogue, data warehouse
 - Researchable (meta) data is made available there. They can be searched and accessed for reuse.
- F4: metadata shall contain data PID
 - Otherwise no way to traverse from metadata to data and its online resources



A

Accessible: Once found the user (humand AND machine) needs to know how to access data

- A1 : (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 : the protocol is open, free, and universally implementable
 - A1.2 : the protocol allows for an authentication and authorization procedure, where necessary
- A2 : metadata are accessible, even when the data are no longer available

- A1: standardized communication protocol
 - Protocol = procedures used by machines to communicate together
 - HTTP and FTP are good examples
- A1.1: open and free protocol
 - Need to be freely usable and interoperable (ex : HTTP, FTP, ...) . Proprietary and closed protocols are to be avoided
- A1.2: support authentication where necessary
 - Accessible does not always imply free and open access all the time. Data can have restricted access for legal reasons (ex : protection of endangered species, national security, ...). Thus communication protocols shall allow to apply restrictions (ex : HTTPS, FTPS, ...)
- A2: metadata are accessible when data is not anymore
 - Metadata allow then to know about the existence of those data and contact responsible parties, find associated scientific articles



Interoperable: to allow data use regardless of the IT environment used

- I1 : (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2 : (meta)data use vocabularies that follow FAIR principles
- I3 : (meta)data include qualified references to other (meta)data

- I1: (meta) data described in a formal, accessible, shared, and broadly applicable language for knowledge representation
 - knowledge representations = data model (UML), ontologies (vocabularies)
 - In 2023, there are numerous international standards and best practices to be reused. They need to be reused.
 - This provides comprehensible resource for human AND machines. Then (meta)data can be exchanged seamlessly (with no human intervention to 'hard-wire/code' things for them to work)
- I2: vocabularies following FAIR principles
 - = vocabularies documented and accessible by a PID, having metadata, being readable by machines
- I3: (meta)data referencing others
 - FAIR principles are highly based on Web of Data standards and best practices
 - Step by step this creates a global network of scientific information linking source (raw) data to others based on them.
 - (meta)data search is then improved and allowing to discover other meaningful data

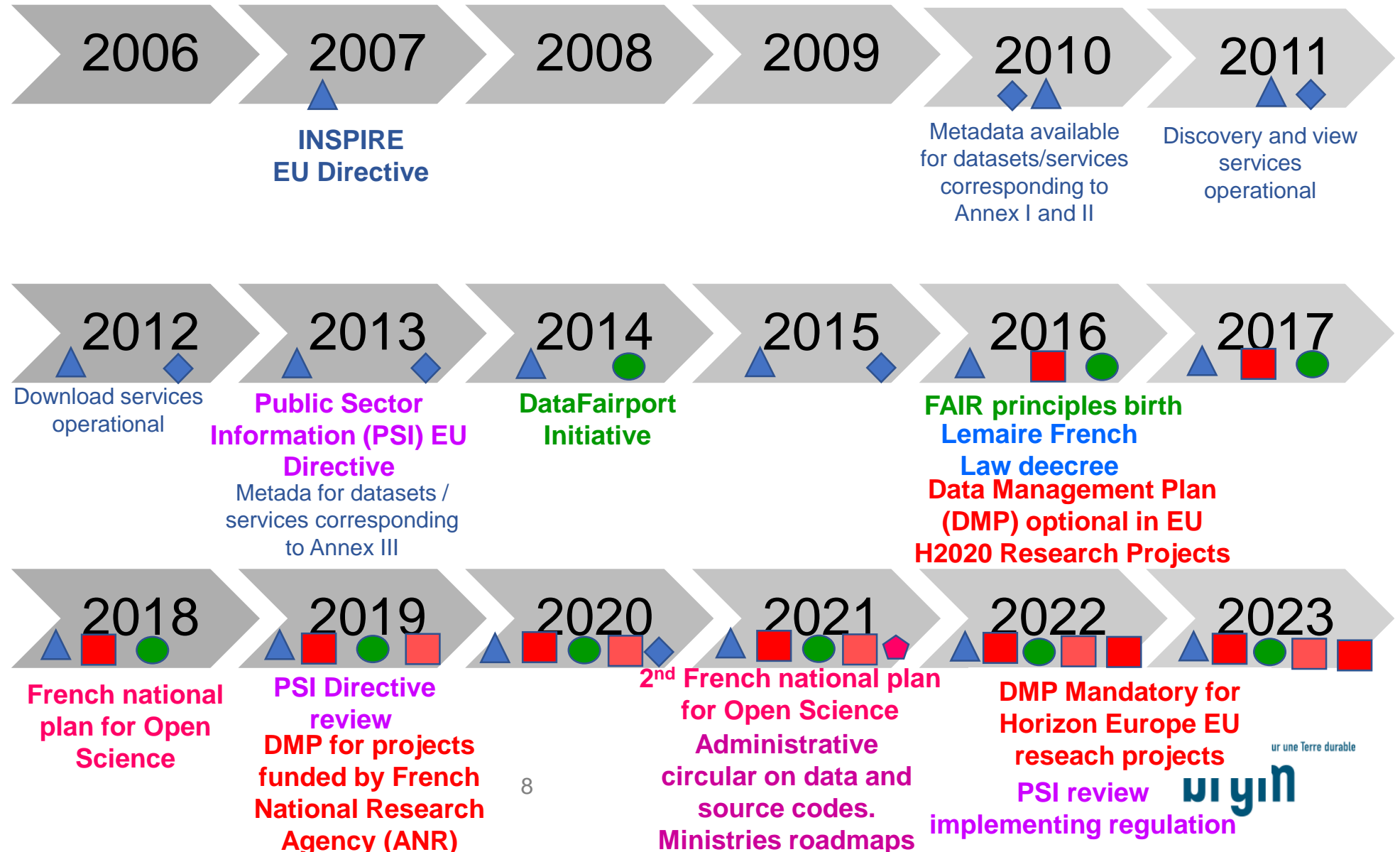


R

Reusable : data need to be reusable for future research

- R1 : (meta)data have a plurality of accurate and relevant attributes
 - R1.1 : (meta)data are released with a clear and accessible data usage license
 - R1.2 : (meta)data are associated with their provenance
 - R1.3 : (meta)data meet domain-relevant community standards
- R1: (meta) data have a plurality of accurate and relevant attributes
 - The more we know the context in which (meta)data was created the more one can understand and reuse them
 - Better to describe (meta)data with a lot of associated attributes to maximize understanding and reuse
- R1.1: clear data usage license
 - Allows to explicitly define usage rights for re-users
 - License choice highly depends on the legal nature of each data (their context of acquisition, etc...)
 - There are free licences adapted for Open Data : [Creative Commons](#), [Open Licence](#), [Open Database License](#)
- R1.2: provenance information
 - Provenance, lineage information is crucial to prove (meta)data are reliable, to assess data quality and allow to reproduce their creation (ex : scientific experience). These have to be provided when data are declared in a searchable resource (like other information : author, institution, dates, ...)
- R1.3: meet domain-relevant community standards
 - Community standards allow to have data following a common model to ease data integration within a community
 - If no community standard exist the best way to share (meta)data should be set up (Re : FAIR principles)

Hindsight on those dynamics





Conclusion

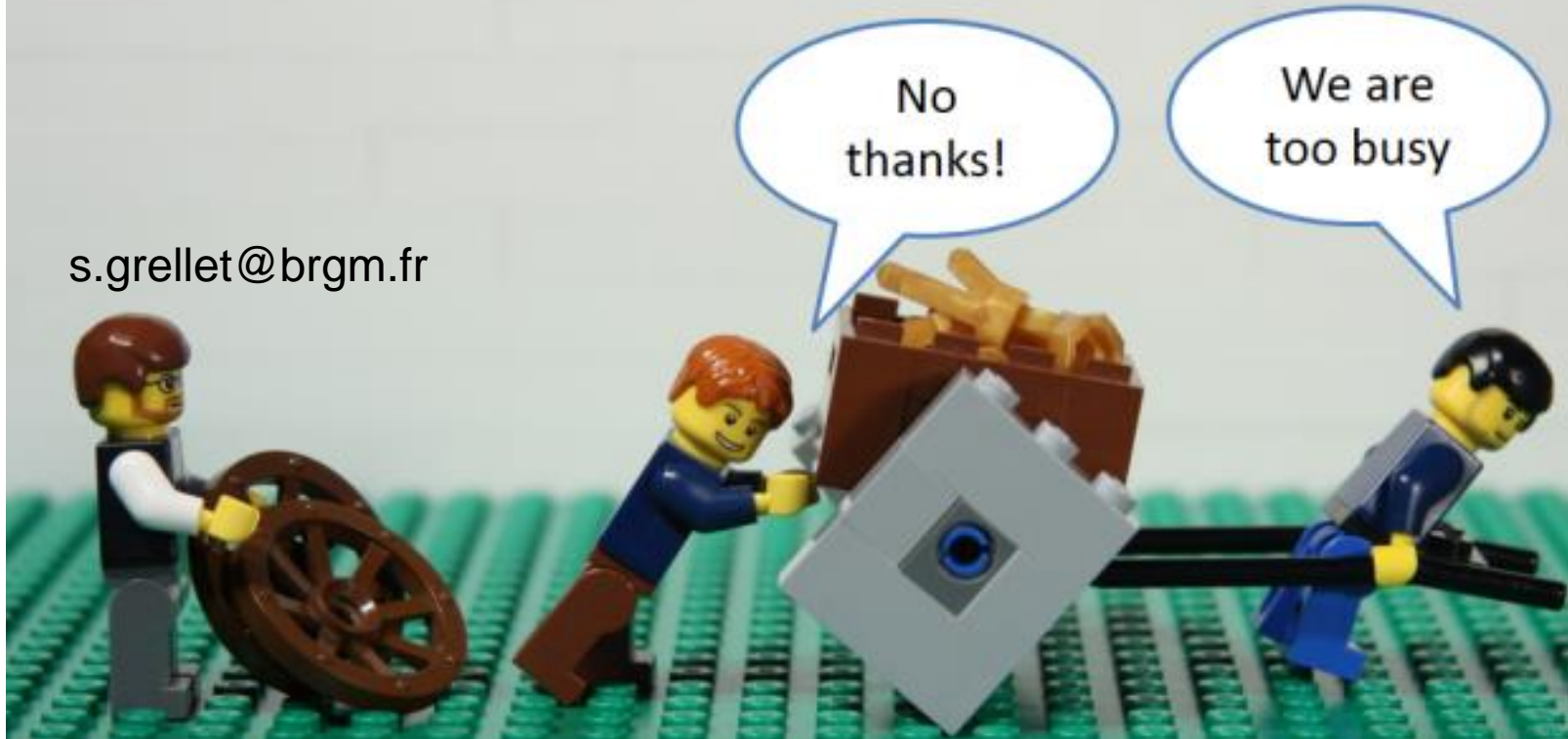
- A continuum towards more open and interoperable access to data
- The research and public domain have to abide by those dynamics
- Those dynamics are highly (if not only) building on internationally agreed up interoperability standards and best practices laid down by community such as W3C, OGC, RDA, INSPIRE
- Exceptions in the law (at least in France) are for some organizations to change their business model that used to be based on selling those data

Thank you



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.performancecompany.com/>



Thanks to

Hélène Bressan

h.bressan@brgm.fr

