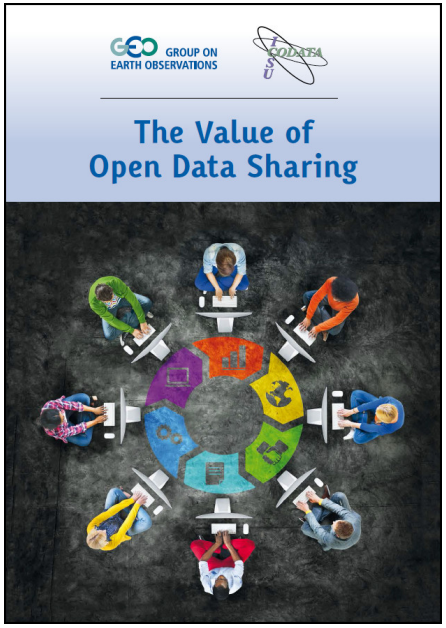
Data Sharing Principles for EO Data

"Focus on GEO and FAIR"

Lionel Menard
lionel.menard@minesparis.psl.eu

Data Sharing Principles - Group on Earth Observation

- **GEO Data Sharing Principles:**



https://doi.org/10.60566/pxdag-hq931

1. Data, metadata and products will be:
   a) **Shared** as **Open Data by default**,
   b) **Following GEOSS Data-CORE license** (Data - Collection of Open Resources for Everyone)

2. Data should be made available**:**
   a) With **minimal restrictions** on use
   b) **At no more** than the **cost of reproduction and distribution**.

3. All data, products and metadata will be:
   a) Made **available** with **minimum time delay**.

How can we implement these sharing principles ?

GEO – Data Management Principles

GEO **Data Management Principles** guidelines* !

- Guidelines include **5 main headings:**
  1. **Discoverability** (Search, catalogues, metadata)
  2. **Accessibility** (On-line services, download)
  3. **Usability** (Standards, documented, traceability, QC,..)
  4. **Preservation** (Protect from loss, integrity, authenticity,…)
  5. **Curation** (Corrections, citations - UUID, DOI)

- Guidelines include **3 levels of conformance**
  1. **Self-assessment & declaration**
  2. **Peer review and automated conformance testing**
  3. **Certification**

DMP label			
	Discoverable	1	D
	Accessible	2	A
	Standard encoding using	3	Usability
	Well documented metadata	4	
	Traceable	5	
	Quality documented	6	
	Preserved	7	Preservation
	Periodically verified	8	
	Reviewed and refreshed	9	Curation
	Tagged with permanent ID	10	

* https://gkhub.earthobservations.org/records/mq2sr-9jp64

FAIR Principles (Findable, Accessible, Interoperable, Reusable)

Published in 2016 in Scientific Data (https://www.nature.com/articles/sdata201618)

- FAIR is targeted for **research data and services**

- **Findable**
  - **Metadata and data** should be **easy to find** for both humans and computers

- **Accessible**
  - Once the **user** finds the required **data**, she/he needs to **know how they can be accessed**, possibly using **authentication and authorization**

- **Interoperable**
  - The **data** usually need to be **integrated with other data**. In addition, the **data need to interoperate** with applications or workflows for analysis, storage, and processing

- **Reusable (Reproducible)**
  - **Metadata and data** should be well-described so that they can be **replicated and/or combined** in different settings

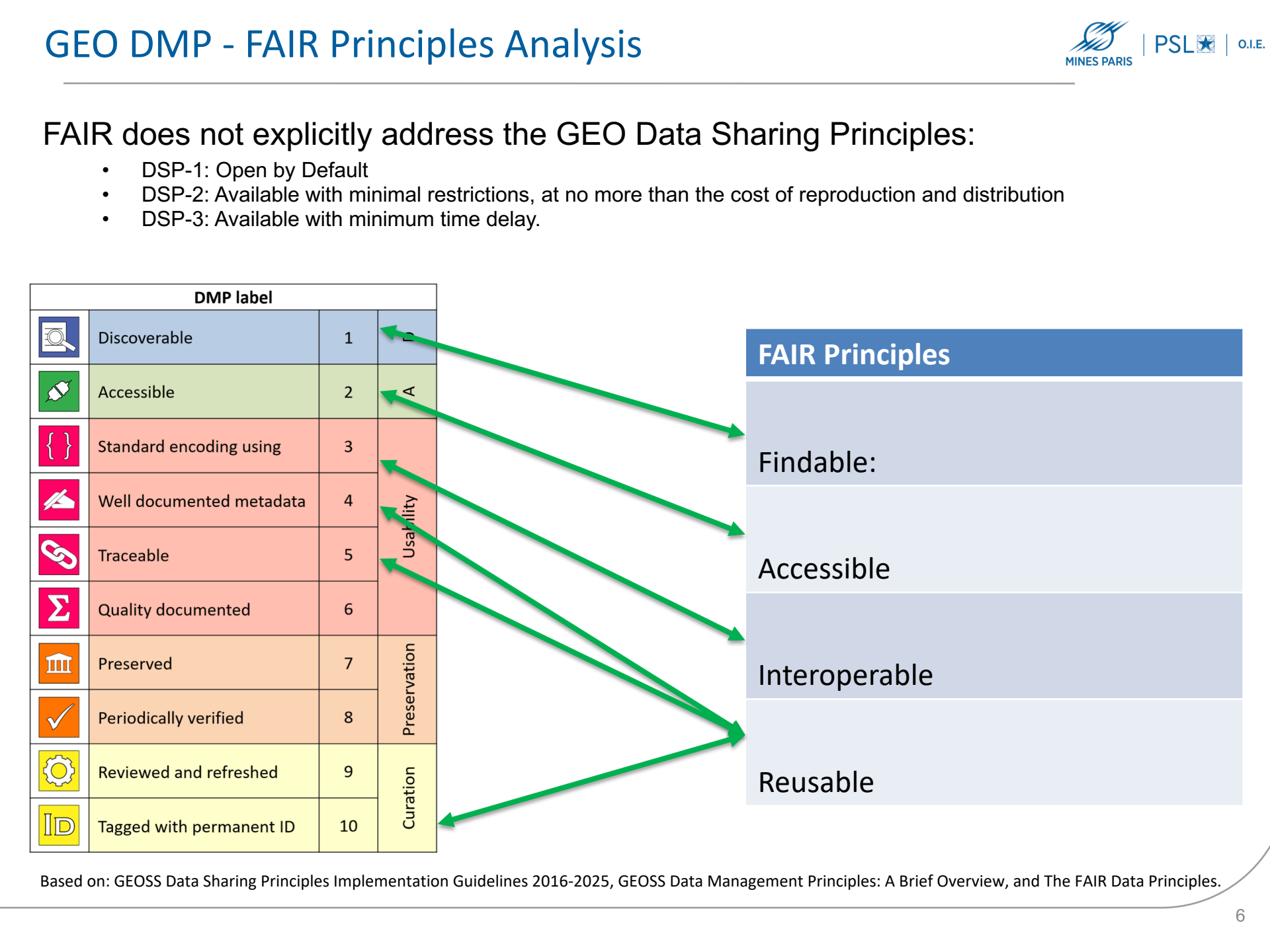
Details of recommendations of Fair Principles: https://www.go-fair.org/fair-principles/

GEO DMP - FAIR Principles Analysis

FAIR does not explicitly address the GEO Data Sharing Principles:

- DSP-1: Open by Default
- DSP-2: Available with minimal restrictions, at no more than the cost of reproduction and distribution
- DSP-3: Available with minimum time delay.

DMP label			
	Discoverable	1	F
	Accessible	2	A
	Standard encoding using	3	Usability
	Well documented metadata	4	
	Traceable	5	
	Quality documented	6	
	Preserved	7	Preservation
	Periodically verified	8	
	Reviewed and refreshed	9	Curation
	Tagged with permanent ID	10	



Based on: GEOSS Data Sharing Principles Implementation Guidelines 2016-2025, GEOSS Data Management Principles: A Brief Overview, and The FAIR Data Principles.

How to report these management principles ?

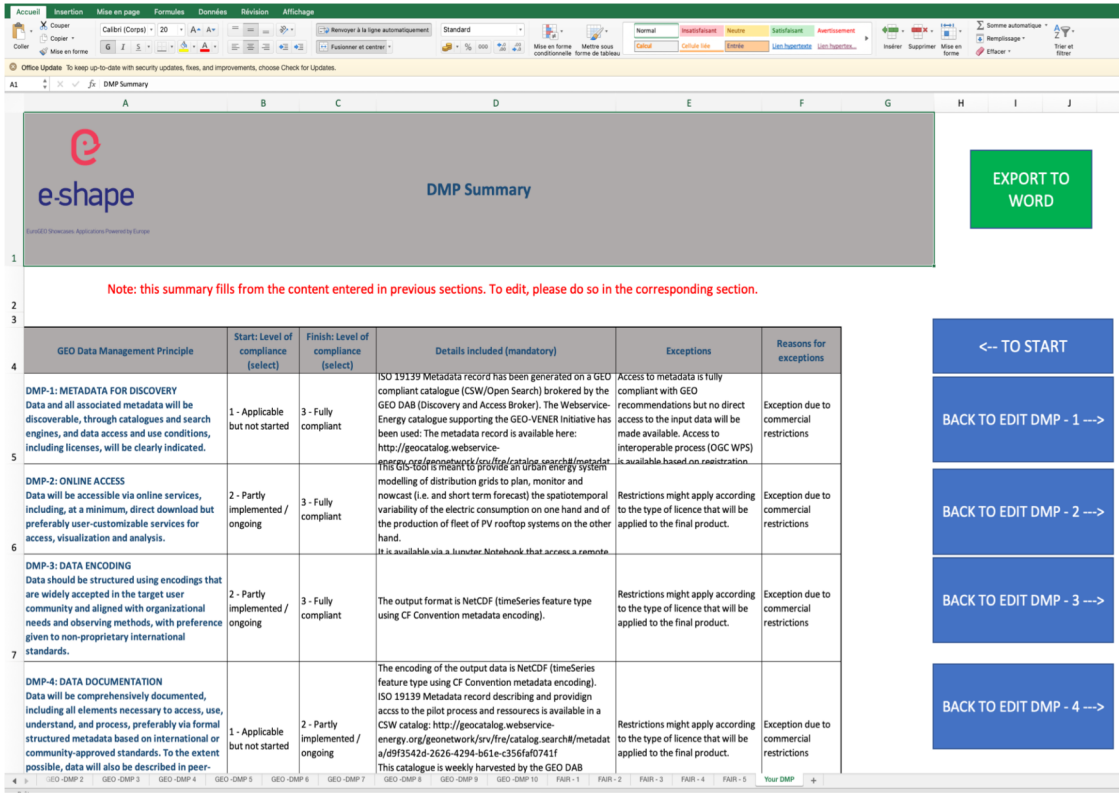
Data management Plan:

- A **living document** outlining **how data** should be **handled** throughout the **lifecycle of a project**
- **Consider the many aspects** of **metadata** and **data:**
  - Generation, management, preservation, storage, dissemination…
- Address **compliance with policies/license**
- Prepare data management for reuse in the **future**

Data Management Plan

GEO Data management self-assessment tool*

- **Designed** for **GEO** and **FAIR** Principles
- **Excel package** (Macro)
- **10 GEO** DMPs and **4 FAIR** Principles
- **Free,** open**, simple** and easy to navigate
- Allow **templating** (Look & feel)
- Based on **self-assessment**
- Provides **recommendation and guidance**
- Notion of **compliance** and **trajectory**



GEO Data Management Principle	Start: Level of compliance (select)	Finish: Level of compliance (select)	Details included (mandatory)	Exceptions	Reasons for exceptions
DMP-1: METADATA FOR DISCOVERY Data and all associated metadata will be discoverable, through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.	1 - Applicable but not started	3 - Fully compliant	ISO 19139 Metadata record has been generated on a GEO compliant catalogue (CSW/Open Search) brokered by the GEO DAB (Discovery and Access Broker). The Webservice-Energy catalogue supporting the GEO-VENER Initiative has been used: The metadata record is available here: http://geocatalog.webservice-energy.org/geonetwork/srv/fre/catalog.search#/metadat	Access to metadata is fully compliant with GEO recommendations but no direct access to the input data will be made available. Access to interoperable process (OGC WPS) is available based on registration	Exception due to commercial restrictions
DMP-2: ONLINE ACCESS Data will be accessible via online services, including, at a minimum, direct download but preferably user-customizable services for access, visualization and analysis.	2 - Partly implemented / ongoing	3 - Fully compliant	This cds-tool is meant to provide an urban energy system modelling of distribution grids to plan, monitor and nowcast (i.e. and short term forecast) the spatiotemporal variability of the electric consumption on one hand and of the production of fleet of PV rooftop systems on the other hand. It is available via a Jupyter Notebook that access a remote	Restrictions might apply according to the type of licence that will be applied to the final product.	Exception due to commercial restrictions
DMP-3: DATA ENCODING Data should be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.	2 - Partly implemented / ongoing	3 - Fully compliant	The output format is NetCDF (timeSeries feature type using CF Convention metadata encoding).	Restrictions might apply according to the type of licence that will be applied to the final product.	Exception due to commercial restrictions
DMP-4: DATA DOCUMENTATION Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-	1 - Applicable but not started	2 - Partly implemented / ongoing	The encoding of the output data is NetCDF (timeSeries feature type using CF Convention metadata encoding). ISO 19139 Metadata record describing and providign accss to the pilot process and ressources is available in a CSW catalog: http://geocatalog.webservice-energy.org/geonetwork/srv/fre/catalog.search#/metadata/d9f3542d-2626-4294-b61e-c356faf0741f This catalogue is weekly harvested by the GEO DAB	Restrictions might apply according to the type of licence that will be applied to the final product.	Exception due to commercial restrictions

- Tested and **validated** on **37 pilots (X2)** of the H2020 EC funded project **e-shape**
- **Open** license and **free** download on the **GEO Knowledge Hub**
  - https://gkhub.earthobservations.org/records/xea2n-tj707

Data Sharing Principles

Key points:

- Data is **key** ! Assessing **compliance to Sharing Principles is crucial**
- Data Sharing/Mangement Principles are:
  - Applicable for **any EO domain**
  - Work for **all type of profiles** and **needs** (Research, education, business,…)
- Reporting tools (**DMP's**) exists
- **Benefits**
  - Improve **data governance** and **data quality**
  - Foster **research** and **innovation**

2024 – Develop a Machine-Actionable GEO Data Management Plan

- **DMP OPIDoR platform**: https://dmp.opidor.fr/
  - Reliable, scalable, interoperable
  - Supports **models**, tools and online services
  - Enables import and export via **API**

Thanks for you attention

Lionel Menard
lionel.menard@minesparis.psl.eu