

# GEO VIRTUAL SYMPOSIUM 2022

GLOBAL ACTION FOR LOCAL IMPACT

## e-shape Data Management Plan Tool A contribution to GEO



e-shape

EuroGEO Showcases: Applications Powered by Europe

May 3rd 2022

**Lionel Menard and Nicolas Fichaux**  
MINES ParisTech, PSL University - ARMINES

**GEO** GROUP ON  
EARTH OBSERVATIONS

#EO4IMPACT



e-shape

EuroGEO Showcases: Applications Powered by Europe

## e-shape - EuroGEO Showcases Applications Powered by Europe

- European contribution to GEO establishing **EuroGEO**
- 15M€, 65 partners, 4 years duration (2019-2023)
- **ARMINES** (France) coordinator – [e-shape.eu](https://e-shape.eu)
- **37 pilots** in **7 showcases** (10 new pilots was on-boarded - 50k€ grant)



agriculture



health



renewable  
energy



ecosystem



water



disaster



climate

# Objectives

- O1: Develop **operational EO services with and for users** active in key societal sectors
- O2: **Demonstrate the benefits of the EO pilots** through the coordinated downstream **exploitation of EO data** and the utilization of existing **EO resources**
- O3: Promote **the uptake of pilots** at national and international scale, across vertical markets (**private and public**) and amongst key user communities
- O4: Enable the **long-term sustainability** of the numerous pilots, their penetration in **public and private markets** and support their upscaling
- O5: Increase uptake by **raising awareness** on the solutions developed through tailored and well-targeted communication, **dissemination and outreach activities**



# Focus - Data Management Plan

- e-shape DMP's associated data policies: **GEO Data Management** and **FAIR principles**
- The **first** phase (month 6) was **based on the initial assessment**:
  - **Limited familiarity of pilots** on the differences and overlaps **between FAIR and GEO DMPs**.
  - **Building comprehension** of the issue through **webinars** (Bob Chen and Bob Downs)
- The **second** phase provided a **tool** (Excel) and a **canvas** to capture:
  - “**Compliance level**” and “**Trajectory**” new notions for both GEO DMPs and FAIR principles
- The **third** phase **assessed, evaluated** and compiled a **comprehensive picture of the status**.
- **Fourth** phase completed by the end of April 2021, **35 DMPs were received (23 did 2 loops)**
  - **GEO DMP**: 69% compliance across the 10 dimensions:
    - 70% for GEO-2, 3, 4, 6, 7 - GEO-1, 5, 8, 9 and 10 (65%, 52%, 62% 62% and 56%).
  - Achieved **FAIR Principles** compliance: F (67%) A (81%) I (59%) R (70%).



# GEO DSP/DMP – FAIR analysis

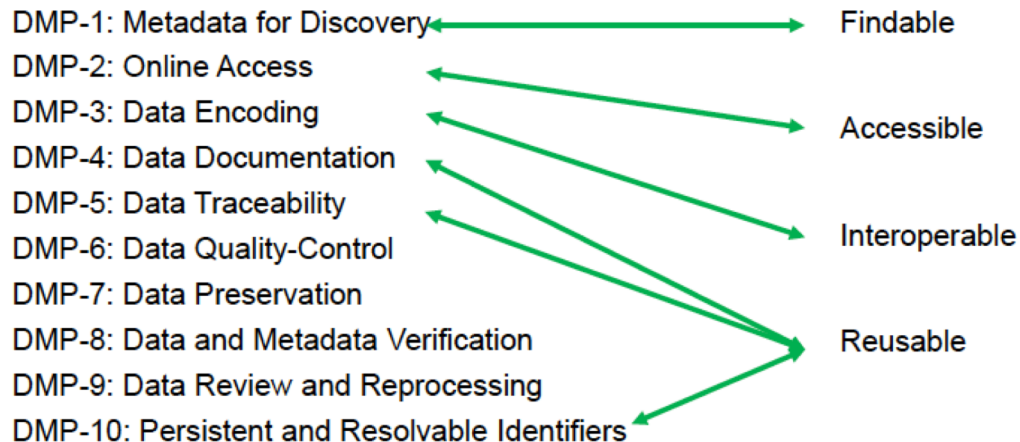


DSP-1: Open by Default

DSP-2: Available with minimal restrictions, at no more than the cost of reproduction and distribution

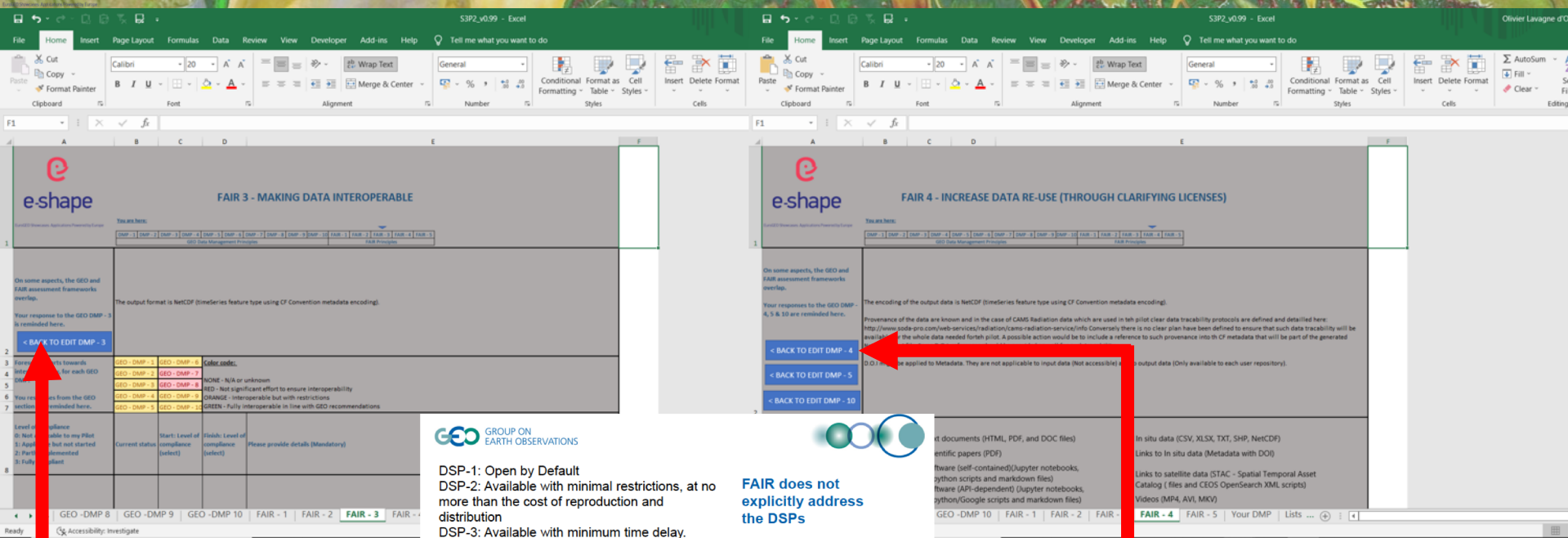
DSP-3: Available with minimum time delay.

**FAIR does not explicitly address the DSPs**



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

# DMP Toolbox (Excel Spreadsheet) addresses GEO/FAIR overlap



**FAIR 3 - MAKING DATA INTEROPERABLE**

**FAIR 4 - INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES)**

**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.**

**FAIR does not explicitly address the DSPs**

**DMP-1: Metadata for Discovery** → Findable  
**DMP-2: Online Access** → Accessible  
**DMP-3: Data Encoding** → Accessible  
**DMP-4: Data Documentation** → Accessible  
**DMP-5: Data Traceability** → Interoperable  
**DMP-6: Data Quality-Control** → Interoperable  
**DMP-7: Data Preservation** → Reusable  
**DMP-8: Data and Metadata Verification** → Reusable  
**DMP-9: Data Review and Reprocessing** → Reusable  
**DMP-10: Persistent and Resolvable Identifiers** → Reusable

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



# DMP Toolbox – Includes 2 additional elements

e-shape requires to **monitor compliance** and **document improvements** during the course of the project (e-shape specific **KPI**)

	Current status	End of Sprint 1 (2021)	End of Sprint 2 (2022)
Level of compliance			
0 : Not applicable to my Pilot			
1: Applicable but not started			
2: Partly implemented			
3: Fully compliant			
Details (mandatory), including if "0"			
Links to knowledge base (if applicable)			

**Compliance level**  
(specific e-shape)

**Trajectory:**  
Introduce a « **forecast** » notion  
(specific e-shape)

# DMP Summary Document including GEO and FAIR sections

S3P2\_v0.99 - Excel

Olivier Lavagne d'Ortigue

File Home Insert Page Layout Formulas Data Review View Developer Add-ins Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

A1 DMP Summary

**e-shape**

**DMP Summary**

EXPORT TO WORD

Note: this summary fills from the content entered in previous sections. To edit, please do so in the corresponding section.

GEO Data Management Principle	Start: Level of compliance (select)	Finish: Level of compliance (select)	Details included (mandatory)	Exceptions	Reasons for exceptions
<b>DMP-1: METADATA FOR DISCOVERY</b> 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 no started	2 - Partly implemented / ongoing	ISO 19115 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 metaata record is available here: <a href="http://geocatalog.webservice-energy.org/openSearch?showMetadataShowDetails&amp;id=731">http://geocatalog.webservice-energy.org/openSearch?showMetadataShowDetails&amp;id=731</a>	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) will be made available based on	Exception due to commercial restrictions
<b>DMP-2: ONLINE ACCESS</b> 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	2 - Partly implemented / ongoing	Invokable on-the-fly by the end users via a Jupyter Notebook like GUI with GIS features. WPS is invokable as well directly for machine-to-machine type of requests. It is available here: <a href="http://s3p2.oie-lab.net/service/wps?Service=WPS&amp;version=1.0.0&amp;request=DescribeProcess&amp;Identifier=s3p2">http://s3p2.oie-lab.net/service/wps?Service=WPS&amp;version=1.0.0&amp;request=DescribeProcess&amp;Identifier=s3p2</a>	Restrictions might apply according to the type of licence that will be applied to the final product.	Exception due to commercial restrictions
<b>DMP-3: DATA ENCODING</b> 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	2 - Partly implemented / ongoing	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

<-- TO START

BACK TO EDIT DMP - 1 -->

BACK TO EDIT DMP - 2 -->

BACK TO EDIT DMP - 3 -->

Ready Accessibility: Investigate



# DMP Summary Document including GEO and FAIR sections

Data Management Plan S3P2.pdf

Ouvrir avec Adobe Acrobat

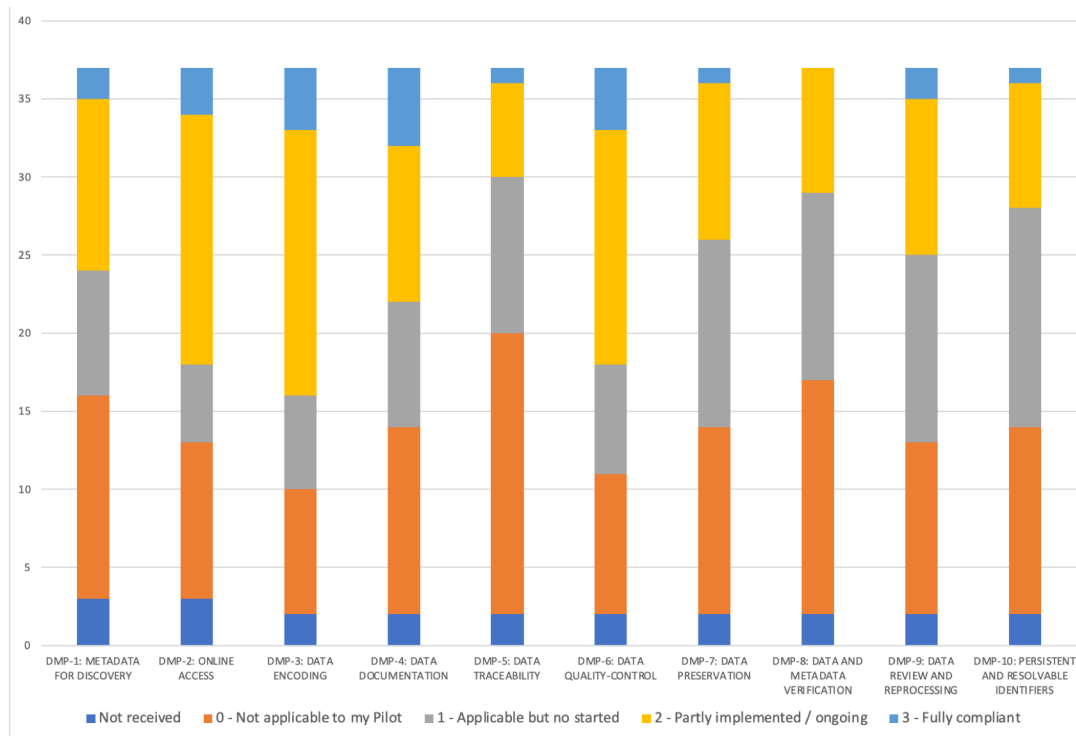


FAIR - Data Management Principle	Start: Level of compliance (select)	Finish: Level of compliance (select)	Please provide details (Mandatory)
<b>FAIR 1 - MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA</b>			
Outline the discoverability of data (metadata provision)	(select value in dropdown menu)	2 - Partly implemented / ongoing	ISO 19139 Metadata via CSW in webservice-energy catalogue the GEO-VENER Initiative support Catalogue: <a href="http://geocatalog.webservice-energy.org/geonetwork/srv/eng/metadata.show?id=8731&amp;currTab=simple">http://geocatalog.webservice-energy.org/geonetwork/srv/eng/metadata.show?id=8731&amp;currTab=simple</a>
Outline the identifiability of data and refer to standard identification mechanism.  Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	1 - Applicable but no started	2 - Partly implemented / ongoing	D.O.I. should be used to refer to the pilot as a whole (eg. As a GEO Knowledge Hub resource).
Outline naming conventions used	1 - Applicable but no started	2 - Partly implemented / ongoing	CF Convention to support the output generated result in NetCDF format
Outline the approach towards search keyword	1 - Applicable but no started	2 - Partly implemented / ongoing	Included into ISO 19139 Metadata record
Outline the approach for clear versioning	1 - Applicable but no started	2 - Partly implemented / ongoing	No clear plan are yer defined
Specify standards for metadata creation (if any).  If there are no standards in your discipline describe what metadata will be created and how.	1 - Applicable but no started	2 - Partly implemented / ongoing	ISO 19139 Metadata via CSW in webservice-energy catalogue the GEO-VENER Initiative support Catalogue: <a href="http://geocatalog.webservice-energy.org/geonetwork/srv/eng/metadata.show?id=8731&amp;currTab=simple">http://geocatalog.webservice-energy.org/geonetwork/srv/eng/metadata.show?id=8731&amp;currTab=simple</a>
<b>FAIR 2 - MAKING DATA OPENLY ACCESSIBLE</b>			



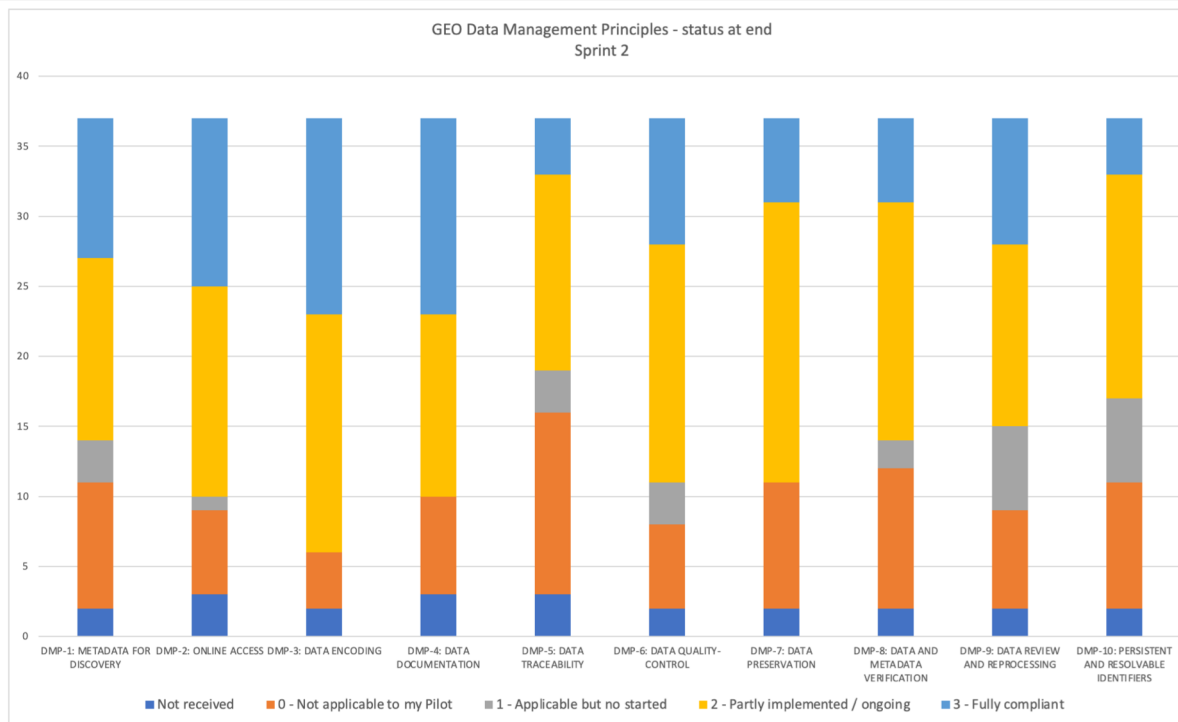
# GEO Data Management Principles

# Status and compliance level at the end of Sprint 1 (April 2021)



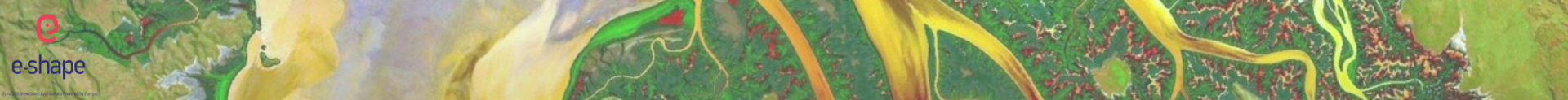
GEO	GEO-1	GEO-2	GEO-3	GEO-4	GEO-5	GEO-6	GEO-7	GEO-8	GEO-9	GEO-10	Avg
Initial	36%	55%	59%	41%	18%	53%	29%	21%	32%	24%	37%

# Status and compliance level at the end of Sprint 2 (April 2022)



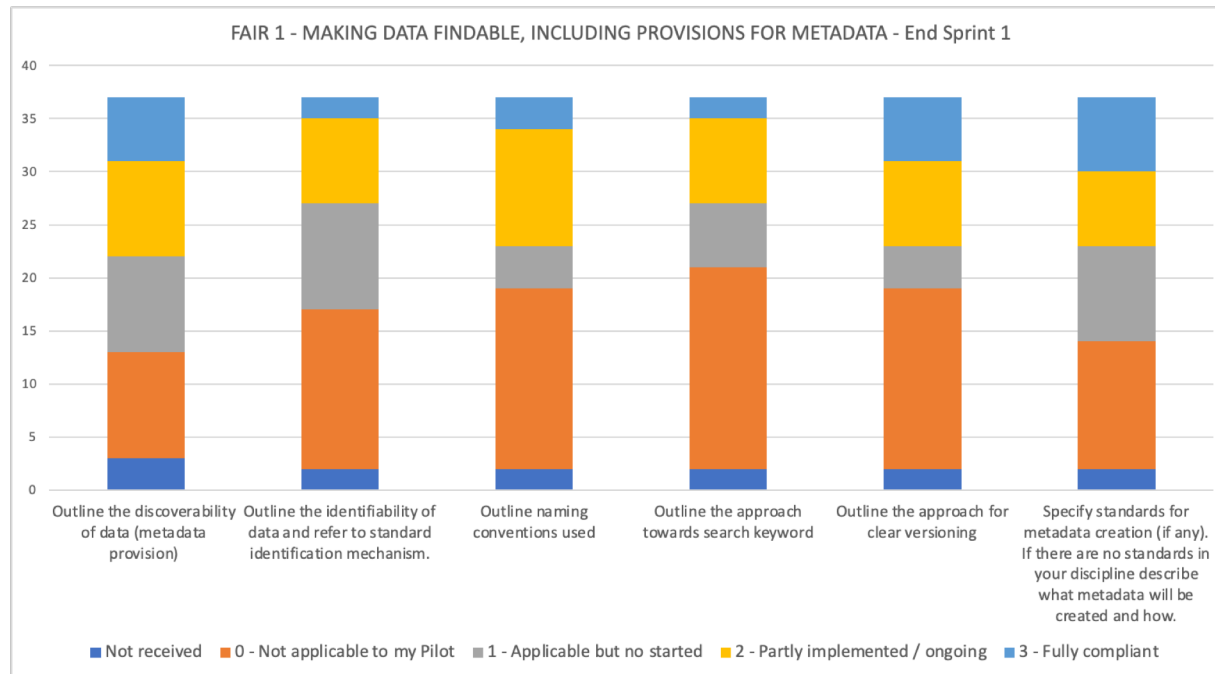
GEO	GEO-1	GEO-2	GEO-3	GEO-4	GEO-5	GEO-6	GEO-7	GEO-8	GEO-9	GEO-10	Avg
Initial	36%	55%	59%	41%	18%	53%	29%	21%	32%	24%	37%
Final	65%	79%	88%	79%	52%	74%	74%	62%	62%	56%	69%





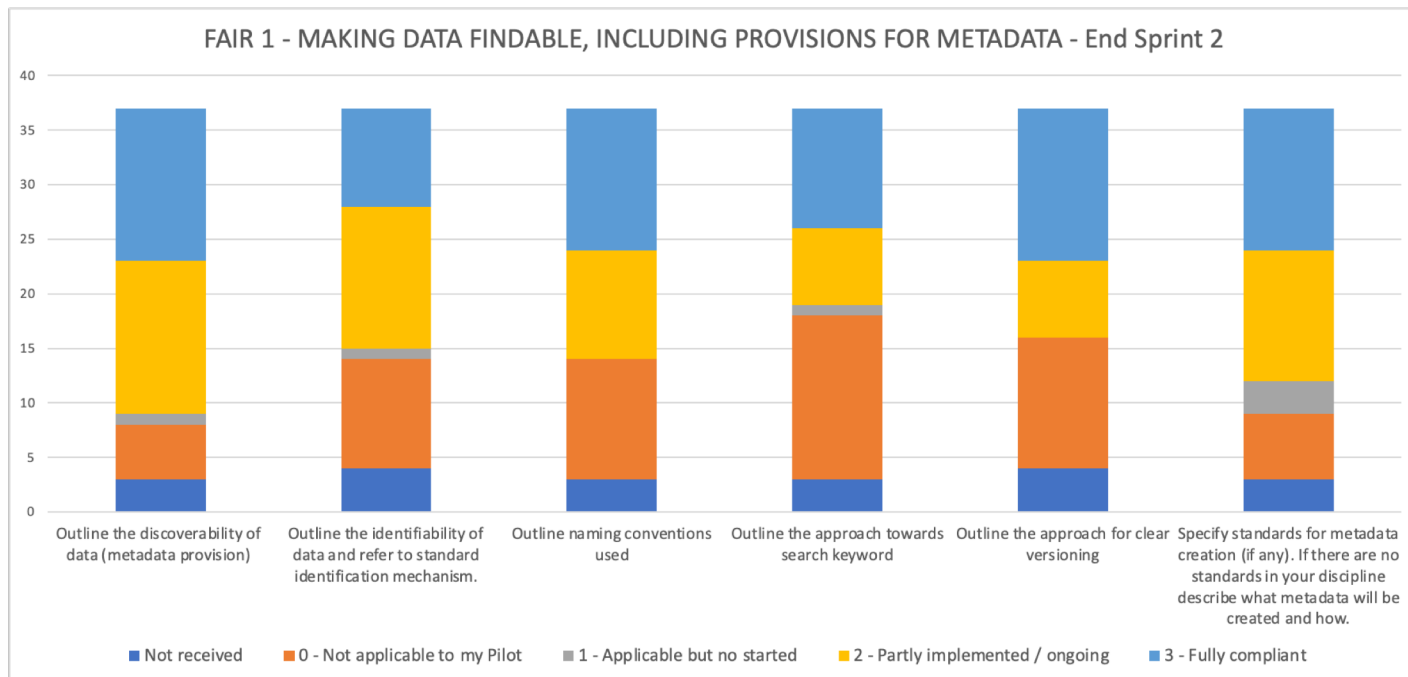
# FAIR Data Management Principles

# Status and compliance level at the end of Sprint 1 (2021)



FAIR 1	FAIR 1-1	FAIR 1-2	FAIR 1-3	FAIR 1-4	FAIR 1-5	FAIR 1-6	Avg
Initial	42%	29%	38%	29%	41%	38%	36%

# Status and compliance level at the end of Sprint 2 (April 2022)



FAIR 1	FAIR 1-1	FAIR 1-2	FAIR 1-3	FAIR 1-4	FAIR 1-5	FAIR 1-6	Avg
Initial	42%	29%	38%	29%	41%	38%	36%
Final	82%	66%	64%	55%	66%	73%	67%

# Lesson learned

From the participation of a **large variety of EO sectors**:

- Good feedback on the **usefulness and simplicity** of the tool (Excel)
- DMP is a very **powerful formal framework** to really and concretely **brainstorm** about your **data** and **services** (Input/output/processes)
- **Data vs services** notions are not straightforward in the guidelines
- In practice it is **not easy to reply to all DMP questions**
- The tool definitely **helps and reward** you with **automatic DMP generation**



# e-shape contributions

- e-shape conducted a 3 years dynamic GEO & FAIR DMP assessments
  - Found the “**Compliance**” and **Trajectory**” concepts really powerful (How to move forward ?)
  - **Polished** and add **genericity** to the tool
  - **Tool is available** for the GEO community (Announcement from GEO web site)
  - On-going work to **move** from Excel towards an **integrated** and **centralized platform**



# Thank You!




EuroGEO Showcases: Applications Powered by Europe

Lionel Menard and Nicolas Fichaux / May 3<sup>rd</sup> 2022  
@eshape\_eu / lionel.menard@mines-paristech.fr

Collaborate and communicate with GEO:





**Lionel Menard** is a research scientist at the Center for Observation, Impacts, Energy at MINES ParisTech / ARMINES acting as the key person for data, results, products dissemination and valorization.

He has an in-depth knowledge of System Development LifeCycle (SDLC), management of Information System and Spatial Data Infrastructure. He has been the team leader of the Energy contribution from phase 2 to 6 of the GEOSS Architecture Implementation Pilots (2006 – 2013). He is a key contributor to the GEO-VENER initiative and the French representative of the GIDTT (GEOSS Infrastructure Development Task Team).

He is the point of contact for OGC at MINES ParisTech. Since 1996, he has been involved in numerous European Commission funded projects playing key role in advocating, designing, prototyping, developing, deploying and monitoring cutting-edge information systems (e-shape (2019-2023), NextGEOSS (2016-2020), ConnectinGEO (2015-2017), EnerGEO (2009-2013), ENDORSE (2011-2013), MESoR (2007-2009), SoDa (2000-2003)).

He is responsible of 2 energy pilots in the framework of NextGEOSS. He is as well involved in the development of one the energy pilots in e-shape while being part of the e-shape project management team (Lead ARMINES). He holds a Master's degree (Ms. Eng.) in Information Systems Management from the University of Nice Sophia Antipolis.



**Lionel Menard**  
Research Scientist  
MINES ParisTech