

# IBERGRID

2024 28-30 OCT  
UNIVERSITY  
OF PORTO

better  
software  
for  
better  
science

13TH IBERIAN GRID CONFERENCE



## Iván Palomo Llavona

**F A I R**eva  
evaluator, validator & advisor

Fair data in the DT\_GEO project



**DT**—**GEO**

# Index

- What is FAIR data?
- What is the FAIR EVA?
- How does the FAIR EVA work?
- How has the FAIR data evolved in the DT-GEO context?

# What is FAIR data?

**F**indable

**A**ccessible

**I**nteroperable

**R**eusable

It refers to three types of entities: data (digital objects), metadata (information about the digital objects) and infrastructure (Where those digital objects live).

Indicators designed by

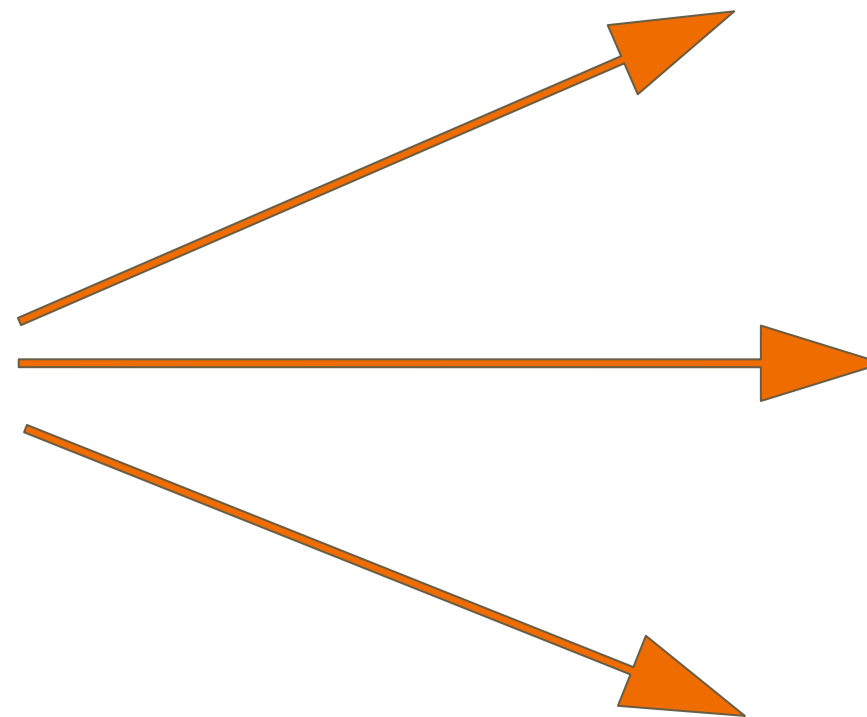


# RDA indicators

- ❑ F1: (Meta)data are assigned globally unique and persistent identifiers.
- ❑ A2: Metadata should be accessible even when the data is no longer available.
- ❑ I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- ❑ R1: (Meta)data are richly described with a plurality of accurate and relevant attributes

# What is the FAIR - EVA

**FAIR-**

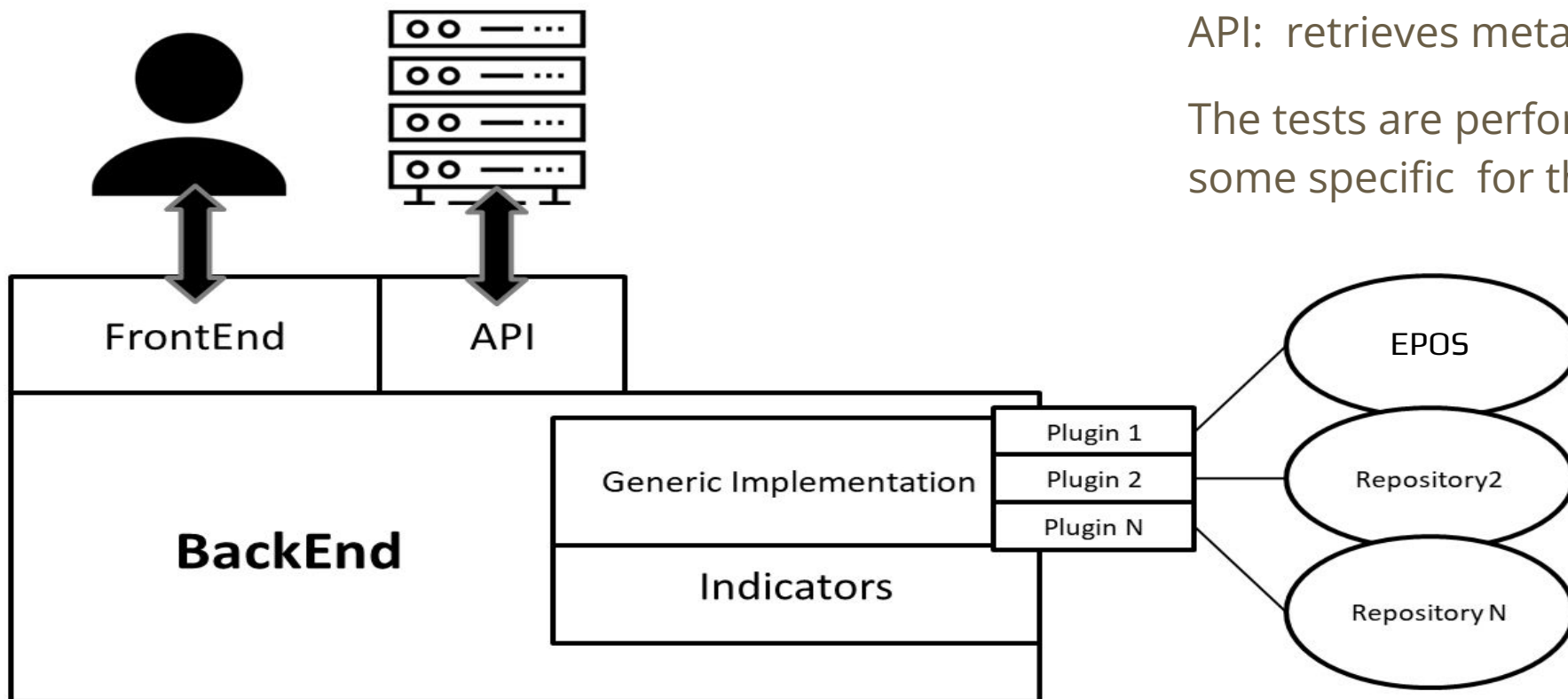


**E**valuator

**V**alidator

**A**dvisor

# How does FAIR-EVA work



User gives: ID and repository.

API: retrieves metadata from repository.

The tests are performed some generic and some specific for the Plugin.



# Example tests

	Indicator	Explanation	Test performed
Findability	RDA-f1-01d	Data is identified by a persistent identifier	Search for identifier and validate it
Accesibility	RDA-a1-01m	Metadata contains information to enable the user to get access to the data	The tool searches for a download URL and the license
Interoperability	RDA-i1-01d	Data use a formal, accessible, shared, and broadly applicable language for knowledge representation	Check that the file & data format in the metadata matches a controlled vocabulary for example: Internet Media Types
Reusability	RDA-r1-01m	(Meta)data are richly described with a plurality of accurate and relevant attributes	Checks the existence of metadata elements related to reusability: formats, license, spatial, temporal

# What does it return

FAIR principle	Score
Findable	66.14
Accessible	88.95
Interoperable	16.93
Reusable	19.12
<b>Total</b>	<b>50.38</b>

ID	Indicator	Score	Output
RDA-F1-01D	Data is identified by a persistent identifier	0	Identifier is not persistent for the data: DT5202
RDA-F1-01M	Metadata is identified by a persistent identifier	100	Found persistent identifier for the metadata: 66d7604d-064f-4e81-a6e2-b539fbb2d91a
RDA-F1-02D	Data is identified by a persistent identifier	0	Identifier found for the data is not globally unique: DT5202
RDA-F1-02M	Metadata is identified by a globally unique identifier	100	Found a globally unique identifier for the metadata: 66d7604d-064f-4e81-a6e2-b539fbb2d91a
RDA-F2-01M	Rich metadata is provided to allow discovery	63	Found 9 (out of 15) metadata elements matching 'Dublin Core Metadata for Resource Discovery' elements
RDA-F3-01M	Metadata includes the identifier for the data	100	Metadata includes identifier/s for the data: ['DT5202']
RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	100	Metadata is gathered programmatically through HTTP (API REST), thus can be harvested and indexed.

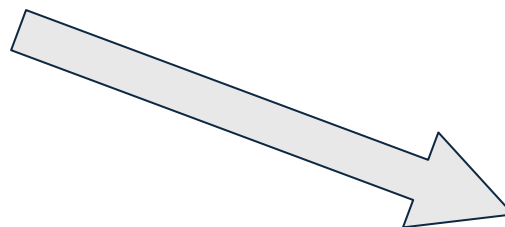


# About DT-GEO



**DT-GEO**

Real time data streams



High precision model of  
the earth

High fidelity models

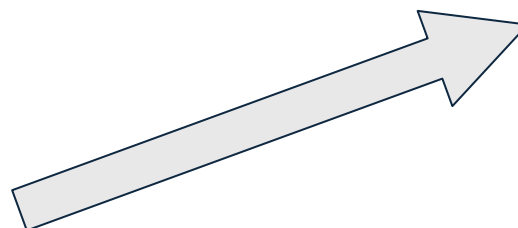
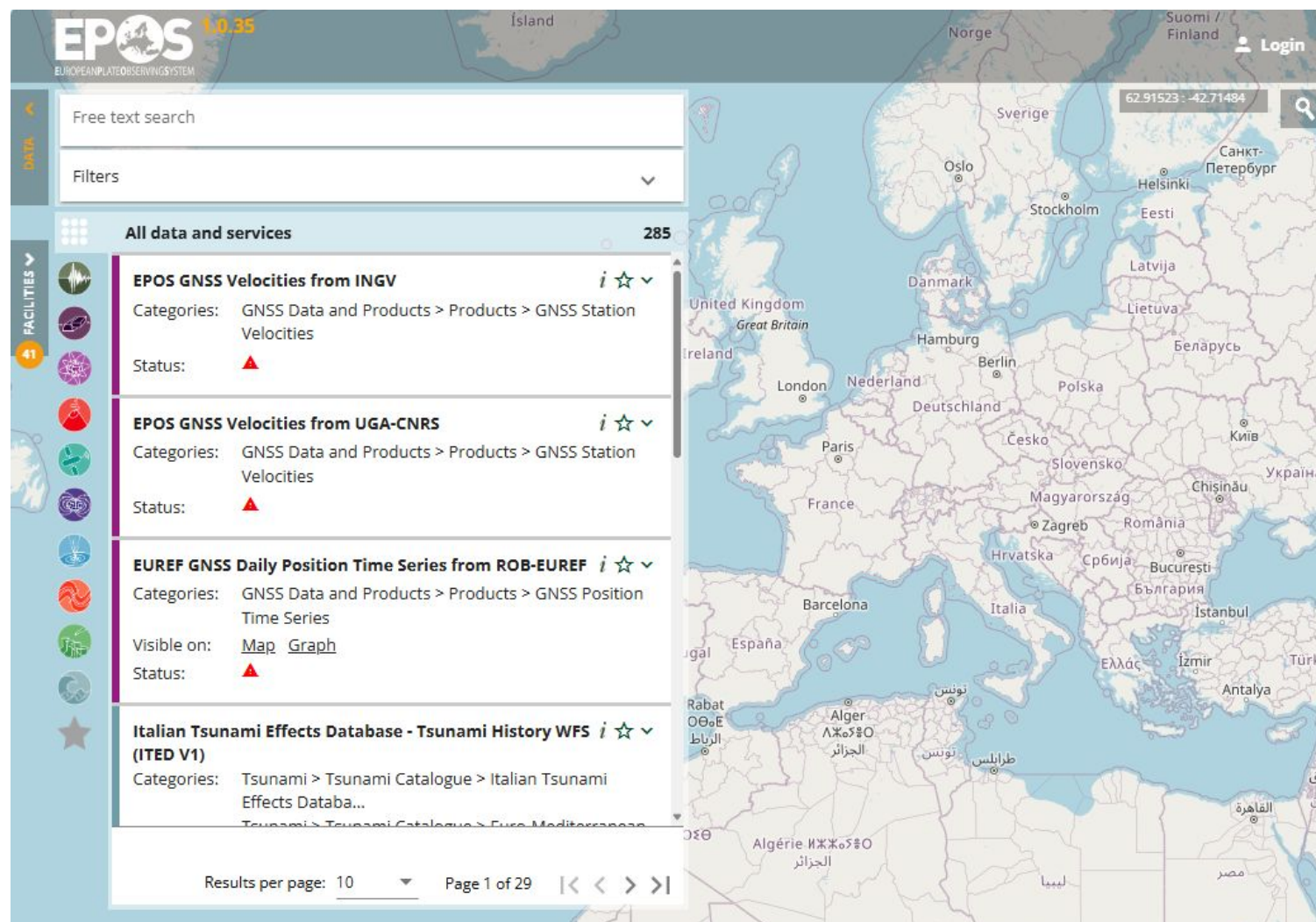


Image of epos data portal

# What are our tools?

We have the the Fair EVA.

And the EPOS Data Catalogue (Infrastructure)



WP 5-8  
Need to be  
characterised

Datasets

Software  
services

Steps

Workflows

Unique ID	Licensing constraints	Person ID	Version	Facility/equipment ID	Security of data transfer
Name	Privacy constraints	Person name	URL	Facility/equipment name	Licensing constraints
Type	Curation and provenance obligations	Person email	Maturity level	Facility/equipment role	Privacy constraints
Keywords	Related DA	Person role	Spatial relevance	Related publications	Curation and provenance obligations
Description	Related DA relationship	Security constraints	Temporal relevance	Quality assurance	Related DA
File format	Attributes and default values	Security of data storage	Organisation	Additional metadata	Related DA relationship

# Curation and vocabularies

Once a lot of the metadata was available we run into some issues analysing the metadata . For example: multiple cases of people using different terms to determine that their object was applicable to not just a specific region of the world.

This required more vocabularies to follow and cooperation from the reviewers to harmonize the metadata.

- “global\_coverage” for digital objects applicable to not just an specific region.
- “infinity” if the item has no explicit end date.
- SPDX is used for licenses
- Internet Media Types for formats

# The prototype data catalogue

In order to adapt to the new metadata categories not included in the EPOS Data catalogue. So a prototype was created in order to have more metadata to get a bigger score

File and data format

Temporal metadata

Organisation and person roles

Provenance



# The EPOS plugin

The generic implementation of FAIR-EVA covers a wide array of repositories, however the creation of a plugin is required in order to work with DT-GEO metadata and the EPOS data catalogue.

- How to get the metadata from the EPOS data catalogue
- Some tests only required to look for the correct category
- Others required to look for more specific data on online platforms like Fairsharing or Internet Media Types.

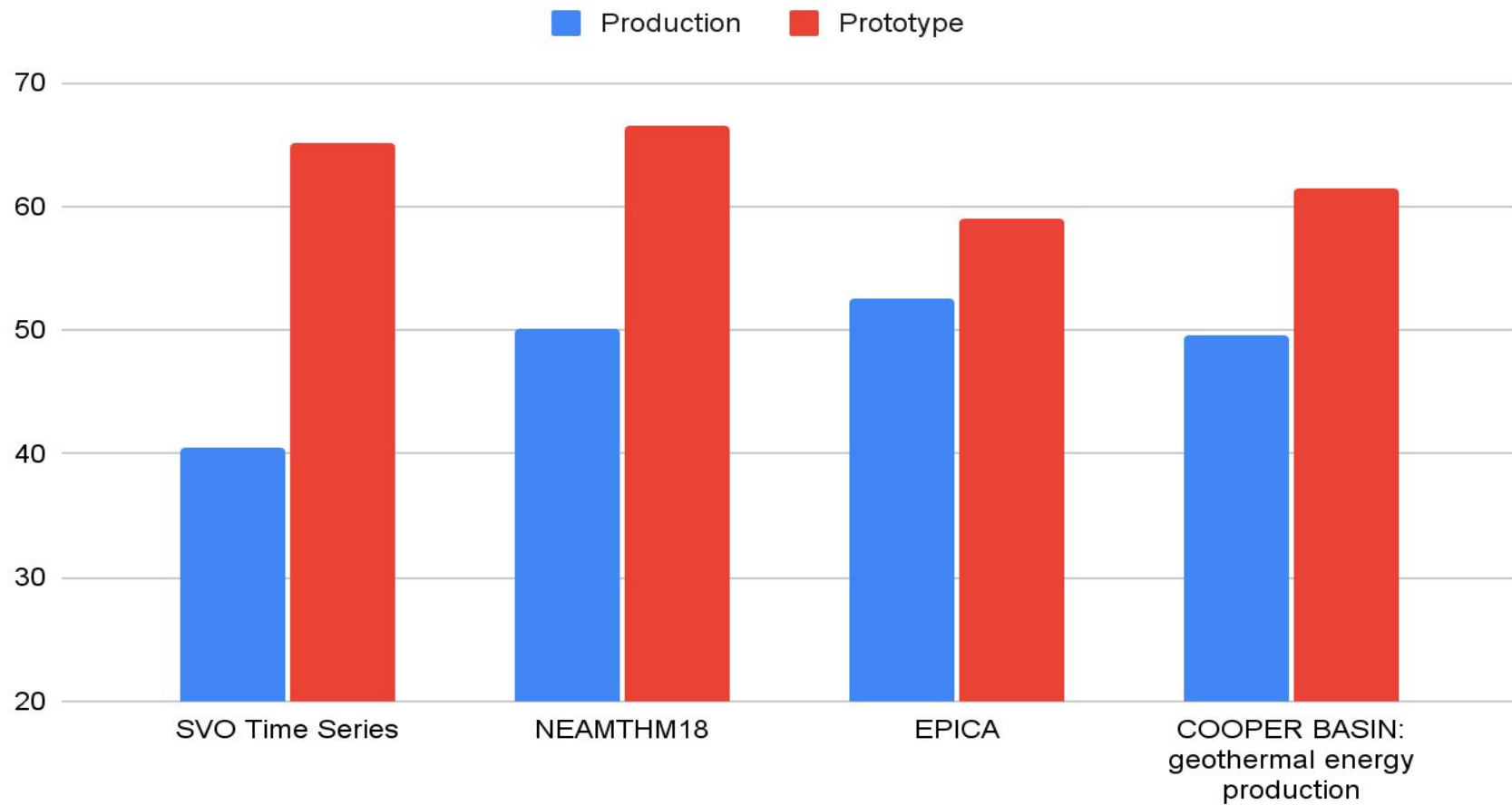
# Tools and utilities developed

Some tools and utilities were developed for the FAIR EVA in order to facilitate its use.

- Its quite an exercise to find the UUIDs (Universally Unique Identifiers) that identify the digital objects in EPOS. So a searcher was added in order to find them by name.
- A python script was created that allows easy access to FAIR-EVA.
- FAIR-EVA was added to SQAaaS platform.
- In order to better manage the evaluation results we added the capability of storing them in csv and feather (fast-on-disk) formats.

# Comparisons before and after

Items  
present in  
production  
and  
prototype  
DTCAT



# Thanks for your attention!



## Any questions?