

FAIRification of geospatial cross-border data



GO-PEG final workshop
28 February 2023

Stefania Morrone (Epsilon Italia)

GOing FAIR with GO-PEG geospatial data



Harmonizing cross-border datasets we:

- were faced with barriers and limitations of non-FAIR data
- worked towards a 'FAIRification' of GO-PEG data
- evaluated the goodness of the results achieved

Here comes an overview of :

- what FAIRness meant to us
- our efforts towards a **FAIRification** of GO-PEG outputs.

INSPIREd FAIRness

Methodology built on INSPIRE, considering its ongoing **simplification and modernization** processes -e.g., Good Practices



Steps towards FAIRification of geospatial data



Findability: facilitated data discovery, especially by a web search engine.

Our data (re)user experience



Findability enablers:

- data catalogs (helping search, discover, understand data assets)
- rich & standard metadata description (easily determine if found asset is suitable for intended use)

Our FAIRification effort

- Create machine-readable metadata (without skimping on detailed descriptions)
- Follow standardized approaches (GeoDCAT-AP, ISO19139)
- Publish MD through National Data Catalogs harvested by the European Data Portal

Steps towards FAIRification of geospatial data



Accessibility: reduced barriers between discovering and accessing data.

Our data (re)user experience



Accessibility enablers:

- clear and simple licensing terms (registration and authentication requirements minimized).
- well-known, open standards and formats for data-sharing

Our FAIRification effort

- Standard licensing terms
- Standard-based data services (OGC OAPIF, WFS...)
- Queries on data enabled, for users to download ‘just what they need’

Steps towards FAIRification of geospatial data



Interoperability: effortlessly and unambiguously exchange data.

Our data (re)user experience



Interoperability enablers:

- PIDs for unambiguous identification of geospatial assets by both humans and machines
- Data specifications
- Agreed-upon controlled vocabularies for unambiguous terms (semantic interoperability)

Our FAIRification effort

- Well-documented data models (mostly INSPIRE-based)
- Code lists registry to enhance semantic interoperability
- Unique identifiers for data and metadata

Steps towards FAIRification of geospatial data



Reusability: facilitate the use of the data or parts of it in different contexts

Our data (re)user experience



Reusability enablers:

- Rich metadata → data quality (accuracy, resolution, completeness, conformity)
- Data Specifications
- Data made available in multiple, standard open formats
- Standard-based services

Our FAIRification effort

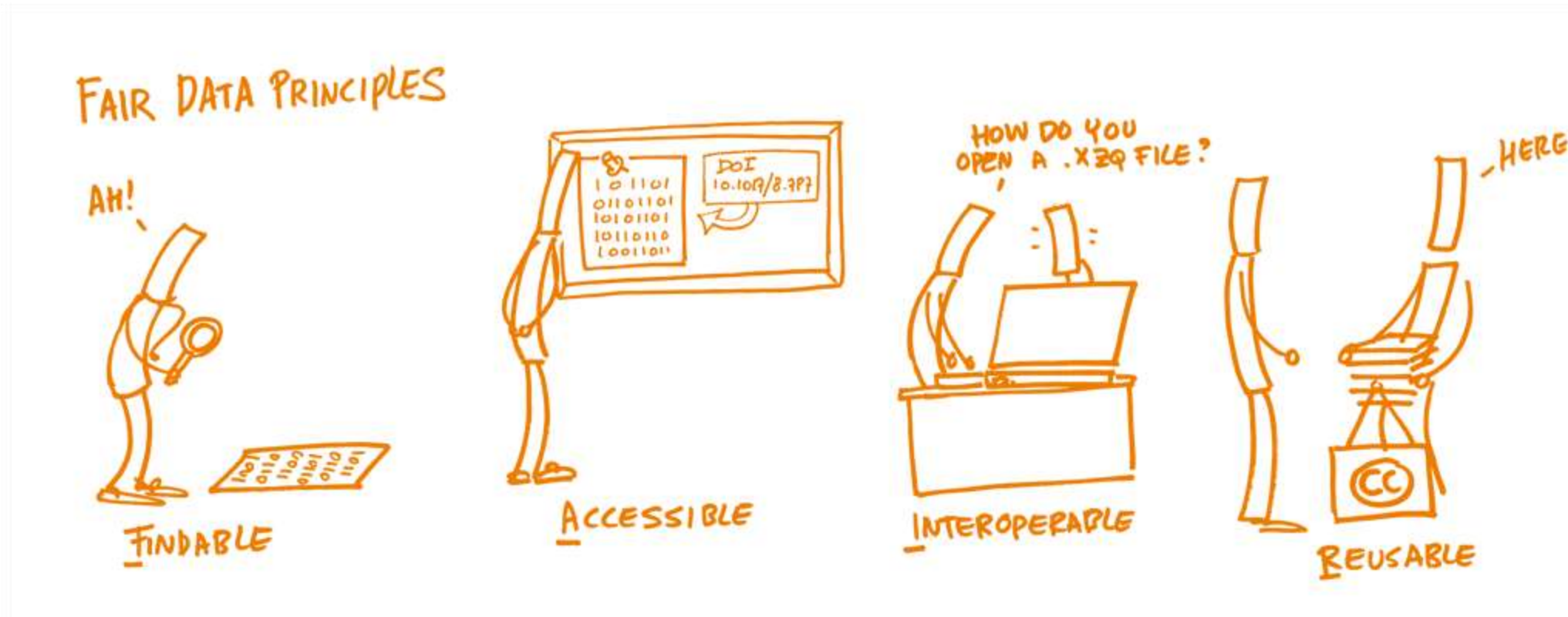
- Rich metadata for data and services, released with a transparent license, adhering to standards and published in EDP.
- Data made available in multiple formats
- Well documented data models: (simplified) INSPIRE data specifications.
- Use of INSPIRE alternative encodings (GeoPackage, GeoJSON)
- Web APIs, namely OGC API services, to make it easier for people and machines to discover, understand and interact with data.

Foster geo-data providers commitment to FAIRness

Alignment with wider frameworks

- FAIRness should not be perceived as an additional burden
- FAIR principles should be embedded in core processes already in place
- National and international initiatives both support and benefit from data FAIRness e.g.,
 - EU data spaces
 - INSPIRE → tangible infrastructure that can support data discovery and sharing.
 - High Value datasets
 - National Data Strategies

Assessing FAIRness of GO-PEG data



Assessing FAIRness : MQA tool

MQA checks

- Compliance with DCAT-AP and DCAT-AP derivatives
- Accessibility of the data referenced in the metadata
- Machine readability of the referenced data
- Use of licenses

“DCAT-AP mandatory fields alone is not sufficient to provide high quality metadata” →
optional fields presence is tested
(and required for 100% validation)

The screenshot shows the official portal for European data, data.europa.eu. The header includes the European Commission logo, a search bar, and links for Login and English. The main navigation bar lists Home, Datasets, Documentation, Publications, data.europa academy, News & events, and Contact us. The breadcrumb trail indicates the path: Home > Metadata Quality Dashboard > Methodology. The page title is 'Metadata quality'. Below the title are three buttons: Dashboard, Catalogues, and Methodology (which is highlighted). To the right of these buttons is a 'Download as report' button with a download icon. The main content area is titled 'Metadata Quality Assessment Methodology' and contains a paragraph explaining the tool's purpose: 'The Metadata Quality Assessment (MQA) is a tool developed by the consortium of data.europa.eu to study the quality of metadata harvested by data.europa.eu. It is intended to help data providers and data portals to check their metadata quality and to receive suggestions for improvements. The results are presented via the MQA and are also available as download. In the following we describe the functionality of the MQA and the methodology it uses.' At the bottom, there is a link to a feedback form: 'If this page still does not answer all your questions, please feel free to contact us via our feedback form at the end of the page.'

Assessing FAIRness

Tools for the automated evaluation of the FAIRness of digital objects

FAIRassist.org



FAIRassist.org

Help you discover resources to measure and improve FAIRness.

FAIRassist is the new, under development, educational component of the well established FAIRsharing resource.

FAIRassist is being designed to offer personalised guidance to all stakeholders to discover standards and repositories in FAIRsharing, which should be used to make data FAIR, as well as a good other resources that enable FAIRness.

FAIRassist is under development, and it will implement a phased rollout of its content, working with and for the community. As an initial step, we are collecting and describing existing resources for the assessment and/or evaluation of digital objects against the FAIR principles, which are aspirational. The focus is on practical questionnaires, checklists and automated tools that help users understand how to achieve a state of "FAIRness", and how this can be measured and improved. This is not intended to be a comprehensive list of all groups, projects and organizations that tackle FAIRness or FAIRification. If your resource is missing, you can submit it via the "Register resource" button below.

[Register a new resource](#)

Resource	Execution Type	Key Features	Organisation	Target Objects	Reading Material
FAIR Data Maturity Tool	Manual - questionnaire	Evaluating systems and reports models	ICIRO-Catania	Datasets	
AutoFAIR	Semi-automated	A tool for automating FAIR assessments for bioinformatics resources	Department of Computer Information Systems, Faculty of ICT, University of Malta	Bioinformatics resources	Published Article
FAIRsharing Checklist	Automated	Checks validity of metadata descriptions, schemes, existence of persistent identifiers (in a context registry) and sometimes of links to data and landing pages.	CLARIN ERIC	Datasets	Documentation
Data Stewardship Wizard	Predictive, Selections manually filled questionnaire	Helps researchers to design a data stewardship strategy for a project aiming for the highest measured FAIR data	ELIXIR NL and ELIXIR CZ	All digital objects	Published Article
Checklist for FAIR	Manual - checklist	A self-assessment tool to measure the FAIRness of an organisation. The self-assessment tool is a simple instrument, presented in an editable PDF form. By answering the questions and evaluating the level (beginner, intermediate, or advanced) at which you assess the performance of your organisation, you will be able to define the actual FAIRness. In addition you can define a Roadmap to become a FAIR Sharing Research Organisation using the information in the more extensive level 3.	ICRDAI Task group: FAIR Sharing Research Organisation	Organisations	
F-GUI	Automated	The F-GUI supports a programmatic assessment of the FAIRness of research data objects based on a set of core metrics developed by the FAIR4HE project. The metrics specification is available at https://doi.org/10.32811/2020-0775758	FAIR4HE (FAIR4HE)	Datasets	Documentation
FAIR Data Maturity Tool	Manual - questionnaire	Evaluating systems and reports models	ICIRO-Catania	Datasets	

Assessing FAIRness: FAIR-Enough tool

Metrics Tests

RESTful API GraphQL API

Evaluate how FAIR a resource is

Provide the URL to the resource, or digital object, you want to evaluate

URL of the resource to evaluate

Search for evaluations

Collections of Metrics Tests

Create a new collection

Access collection	Title	Homepage	Date cre...	Author
fair-enough-metadata	FAIR Enough metadata maturity indicators	https://github.com/MaastrichtU-IDS/fair-enough-m	2022-05-13	https://orcid.org/0000-0002-1501-1082
fair-evaluator-maturity-indicators	FAIR Evaluator maturity indicators	https://github.com/FAIRMetrics/Metrics	2022-05-13	https://orcid.org/0000-0001-6960-357X
fair-enough-data	FAIR Enough data maturity indicators	https://github.com/MaastrichtU-IDS/fair-enough-m	2022-05-13	https://orcid.org/0000-0002-1501-1082
rare-disease-maturity-indicators	FAIR maturity indicators for Rare Disease	https://github.com/LUMC-BioSemantics/RD-FAIRm	2022-05-13	https://orcid.org/0000-0002-1501-1082

Not applicable to GO-PEG geodata context

FAIR-Enough metadata collection metrics

<https://fair-enough.semanticscience.org/collections/fair-enough-metadata>

Identifier	Name
F1_1M	Resource identifier is persistent
F1_2M	Resource metadata identifier is unique
F2_1M	Metadata is grounded and machine-readable
F2_2M	Metadata is structured
F3_1M	Metadata identifier explicitly in metadata
F4_1M	The resource is indexed in a searchable resource
A1_1M	Metadata uses an open free protocol for metadata retrieval
A2_1M	Metadata is persistent
I1_1M	Metadata uses a formal structured knowledge representation language
I1_2M	Metadata uses a formal semantic knowledge representation language
I2_1M	Metadata uses FAIR Vocabularies registered in known repositories
I2_2M	Metadata uses resolvable FAIR Vocabularies
I3_1M	Metadata contains outward references
R1_1M	Metadata includes a License
R1_2M	Metadata includes a standard License

FAIR-ENOUGH vs GO-PEG FAIRness KPIs

FAIR-Enough tests not applicable to GO-PEG	
Accessible	A2 - Metadata is persistent

Tests if metadata contains a persistence police, i.e., if metadata is accessible even when the data is no longer available

GO-PEG KPIs not included in FAIR-ENOUGH		
Accessible	GOPEG-A2	Stability of Access
Accessible	GOPEG-A3	Granularity of Access
Interoperable	GoPEG-I-03D	Data conformance
Reusable	GOPEG-R-01	Rich assets description
Reusable	GOPEG-R-3	Asset lineage



The FAIR-enough assessment is not focused on geospatial data. There's a need for fine-tuning...

Need for harmonised (geo)FAIRness assessment

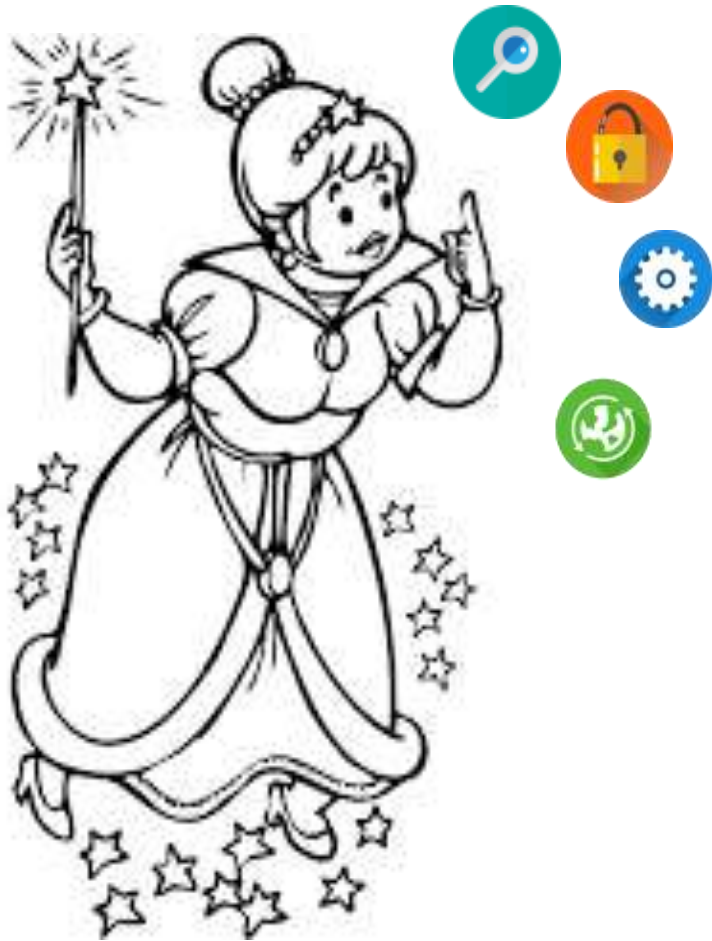
	ITEM UNDER TEST	VALIDATION TOOL			
		INSPIRE Validator	GEODCAT-AP Validator	FAIR-Enough	MQA Tool
<div> <div>Converter 1</div> <div>Converter 2</div> </div>	xml metadata ISO-19139 published in national geoportal	✓			
	rdf metadata (geo)DCAT-AP by data.europa.eu (geo)		✗	6/16	34 violations 5 warnings
	rdf metadata / transform open data port		✗	4/16	30 violations 2 warnings

Assessments results are influenced by conversion processes

Harmonisation needs governance !

FAIR related, but using different metrics

Conclusions



FAIRness adds value to geospatial data, especially when it's cross-border

...But there is no magic to make data FAIR

- FAIRness assessment needs customization to geospatial and harmonisation (**Key FAIRness indicators**)
- INSPIRE (modernized/simplified) is a solid building base, not only for the geospatial domain
- FAIRness assessment results should not be seen as Boolean values (True/False) but rather as ranges or spectrum
- Alignment with EU Data Policy Context can improve data providers commitment

THANK YOU