

Automatic evaluation of FAIR metrics

Alban Gaignard, Thomas Rosnet, Frederic de Lamotte, Marie-Dominique Devignes,

(ELIXIR-France interoperability working group)





Context: FAIR principles



https://www.go-fair.org/fair-principles https://www.nature.com/articles/sdata201618

#ELIXIR21



FAIR principles

- critical for open & reproducible sciences
- result in many guidelines
- technology agnostic guidelines

How to implement the principles and go beyond checklists ?

Resource providers and developers need help, and tooling !



Motivating use cases

- I'm a **data producer**, I published my dataset through an online registry, does it provides rich metadata ?
 - Are these metadata interoperable, reusable ?
 - Is the registry exposing metadata through a community agreed controlled vocabulary ?
- I'm a **software developer**, my source code is on GitHub, but not mature enough to be part yet of a registry.
 - How to check if my tool is **findable** enough ?
 - Which kind of metadata should I advertise ?
 - Am I missing **required** or **recommended** metadata ?







Already running tools

FAIRassist.org \rightarrow Automated tools : 3 out of 13 initiatives



https://fairsharing.github.io/FAIR-Evaluator-FrontEnd



Disclaimer: The test results shown here are based on preliminary data and code which still is under development. F-UJI is rapidly evolving and not yet available in a productive environment.

Research Data Object (URL/PID):*	OAI-PMH:	
http://bio.tools/jaspar	(Optional) Enter the URL of an OAI-PMH endpoint s	run test
Enable caching? (2) 💟 Use DataCite? (2)		

Assessment Results:

https://www.f-uji.net/index.php







Why another tool ? Objectives ?

Assumptions

• Linked Data principles and Semantic Web technologies are at the core of most of the FAIR principles (especially F, I, and R) ... but technical skills are needed.

Aims

- Propose a (simple) web interface for developers to interact with "FAIR Evaluation DEMO Services" (APIs) and make progress on FAIRification (iterative testing)
- Leverage semantic web technologies (RDF, SPARQL, SHACL) to enhance the quality of metadata, with a focus on Findability, and Reusability principles
- Share implementations, and contribute to larger FAIR evaluation initiatives



DEMO







Ontology terms already known ?

F4: (Meta)data are registered or indexed in a searchable resource I2: (Meta)data use vocabularies that follow the FAIR principles R1.3: (Meta)data meet domain-relevant community standards

Metadata may rely on semantic types and relations defined through controlled vocabularies or formal ontologies.

Are these types and relations registered in reference ontology catalogues?

FAIR-checker ask online registries to check that ontologies / vocabularies are "recognized":

- LOV, Linked Open Vocabularies (general purpose, SPARQL endpoint)
- **OLS, Ontology Lookup Service** (life sciences, REST API)
- [planned] BioPortal (life sciences) SPARQL + REST, access controlled (API-key)







3 - Ontology checks

Check Vocabularies

LOV

Classes

http://schema.org/SoftwareApplication

Demo this afternoon !

Properties

http://ogp.me/ns#description	LOV
http://ogp.me/ns#image	LOV
http://ogp.me/ns#title	LOV
http://schema.org/additionalType	LOV
http://schema.org/citation	LOV
http://schema.org/contributor	LOV
http://schema.org/description	LOV
http://schema.org/license	LOV
http://schema.org/name	LOV
http://schema.org/operatingSystem	LOV
http://schema.org/url	LOV
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	LOV
http://www.w3.org/1999/xhtml/vocab#role	unknown
https://bio.tools/ontology/primaryContact	unknown







Metadata quality ("completeness")

R1.3: (Meta)data meet domain-relevant community standards

SHACL : RDF Shape Constraint Language

- Declare constraints / patterns for knowledge graphs
- Matched against **specific** resources (*e.g.* tools, workflows, datasets, articles, etc.)

FAIR-Checker transforms (template engine) Bioschemas profiles into SHACL shapes

Results of shape validation constitute a knowledge graph:

- Filter severity (**warnings** or **errors**)
- Produce user-oriented messages and provide
- #ELIXIRzecommendations





Bioschemas is a community effort aimed at reusing and extending Schema.org for better life science digital resource findability. Several profiles are defined for each kind of Life Science resources, specifying minimal, recommended or optional information. Am I missing minimal information ? Should I provide other information for better findability ?

At the moment, profiles supported are: ScholarlyArticle, Dataset, ComputationalTool

Check BioSchemas

Requirements	Improvements
Property http://schema.org/identifier must be provided	Property http://schema.org/backstory should be provided
Property http://schema.org/headline must be provided	Property http://schema.org/alternateName should be provided
Demo this afternoon !	Property http://schema.org/pageEnd should be provided
	Property http://schema.org/citation should be provided
	Property http://schema.org/pageStart should be provided
	Property http://schema.org/dateModified should be provided
	Property http://schema.org/about should be provided
	Property http://schema.org/isBasedOn should be provided
	Property http://schema.org/license should be provided
	Property http://schema.org/isPartOf should be provided
	Property http://schema.org/dateCreated should be provided
Annotate missing B	ioSchemas properties





Usage statistics



Global indicators : counts of success and failures per week

We would all like to see more **green** bars growing ...







Conclusion and perspectives

A web tool - under active development - to

- leverage FAIR community metrics (FAIR Evaluation Services APIs)
- inspect semantic web standards, to enhance metadata quality (*F*, *I* and *R*)
- provide developers with recommendations and tools for FAIRification

Current limitations

#FI IXIR21

- Early prototype: UI to be improved
- Not yet supporting
 - all Bioschemas profiles
 - BioPortal for Life Science ontologies
- May face "quota" from external resources (WikiData, DataCite)

Several opportunities

- Prototype and share new assessments
- Supporting other Bioschemas profiles
- Supporting other guidelines (RDA Maturity Indicators / EOSC)
- Linking with the FAIR cookbook
- Exporting results / visualization / API







Acknowledgments

ELIXIR-FR interop' working group

Mark D. Wilkinson: using the FAIRmetrics APIs

Funding:

Contacts alban.gaignard@univ-nantes.fr marie-dominique.devignes@loria.fr thomas.rosnet@france-bioinformatique.fr

Live demo: https://fair-checker.france-bioinformatique.fr GitHub: <u>https://github.com/IFB-ElixirFr/fair-checker</u> Issues: https://github.com/IFB-ElixirFr/fair-checker/issues









