FAIR Characterization of Materials Cloud

Leonardo Guerreiro Azevedo, Julio Tesolin, Gabriel Banaggia
IBM Research – Brazil
lga@br.ibm.com, {julio.tesolin,gbanaggia}@ibm.com

Abstract

This document presents the FAIR characterization of digital objects of Materials Cloud using our FAIR questionnaire which is responded employing the Materials Cloud's FAIR Implementation Profile (FIP).

1 The Characterization

Materials Cloud provides an ecosystem that supports researchers throughout the life cycle of a scientific project, and helps them make their researcher output FAIR and reproducible [4]. It uses Invenio framework [1] to build its data repository.

This document presents the answers for our proposal of FAIR questionnaire (version 23.02.2023) considering the responses of the Material Cloud's [3] for the GO FAIR FIP (FAIR Implementation Profile [2]). Table 1 presents the context characterization. Table 2 and Table 3 depict the responses for questions about Findability principles. Table 4 shows the answers for Accessibility principles. Table 5 and Table 6 presents the responses for Interoperability principles. Finally, Table 7 and Table 8 present the answers for Reusability principles.

Table 1: Questionnaire answers for the context of the FAIRness assessment of Materials Cloud.

Q.ID	Questions
Q1	What is your community? Materials Cloud [4].
Q2	Which digital object will be evaluated in this assessment? This assessment is answered using the Materials Cloud FIP [3] which does not make explicit which digital object is assessed.

References

[1] CERN. Invenio digital library framework. https://invenio.readthedocs.io/en/maint-3.1/general/introduction.html, 2015. Accessed in 2023–02-15.

[2] GO FAIR Initiative. Fair implementation profile (fip) mini-questionnaire. https://bit.ly/yourFIP. Accessed in 2023–02-02.

[3] Valeria Granata. Materials cloud fip. https://archive.materialscloud.org/static/documents/fip/Materials%20Cloud%20V1.1.pdf, 2020.

[4] Leopold Talirz, Snehal Kumbhar, Elsa Passaro, Aliaksandr V. Yakutovich, Valeria Granata, Fernando Gargiulo, Marco Borelli, Martin Uhrin, Sebastiaan P. Huber, Spyros Zoupanos, Carl S.

Table 2: Questionnaire answers to assess F1 considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
F1	Q3	What is the attribute that identifies the data (*i.e.*, data is understood as any digital object)? Use of Universally Unique Identifier generated by Invenio 3 framework. It is intended to move to Handle System planned for proper GUPRIs (Globally Unique, Persistent and Resolvable Identifiers) for datasets. .
	Q4	Is the data identifier (ID) globally unique or is it only unique in the database domain or for a specific context? Currently, it is used a Universally Unique Identifier generated by Invenio 3 framework which is unique in the repository. However, they plan to use a Handle System in the future for handling GUPRIs.
	Q5	Is the data ID persistent? It is not possible to answer this question based on the Materials Cloud's FIP.
	Q6	Is the data ID resolvable, *e.g.*, to a landing page? It is not possible to answer this question based on the Materials Cloud's FIP.
	Q7	Are there metadata describing the data? Yes, there are.
	Q8	Do the metadata have a distinct ID from the data? Yes, data and metadata have distinct identifiers. *If the data and metadata have distinct IDs*:
	Q9	- What is the identifier of the metadata? Digital Object Identifier (DOI).
	Q10	- Is the metadata ID globally unique or is it only unique in the database domain or for a specific context? It is globally unique.
	Q11	- Is the metadata ID persistent? Yes, it is persistent.
	Q12	- Is the metadata ID resolvable, *e.g.*, to a landing page? Yes, it is resolvable to a landing page.

Adorf, Casper Welzel Andersen, Ole Schütt, Carlo A. Pignedoli, Daniele Passerone, Joost VandeVondele, Thomas C. Schulthess, Berend Smit, Giovanni Pizzi, and Nicola Marzari. Materials Cloud, a platform for open computational science. *Scientific Data*, 7(1):299, September 2020. Number: 1 Publisher: Nature Publishing Group.

[5] wikidata. Open access. `https://www.wikidata.org/wiki/Q232932`. Accessed in 2023–02-28.

Table 3: Questionnaire answers to assess F2, F3 and F4 considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
F2	Q13	Which metadata schemas, if any, are used to describe the data? Dublin Core (http://www.wikidata.org/entity/Q624610), Schema.org (http://www.wikidata.org/entity/Q3475322), DataCite (http://www.wikidata.org/entity/Q821542)
	Q14	What kinds of metadata (*e.g.*, descriptive, administrative and structural) are used to describe the data? Not answered considering the FIP.
	Q15	Which of these metadata schemas are domain specific and which are domain-agnostic? Not answered considering the FIP.
F3	Q16	What is the technology that links metadata to the data (and vice-versa)? Built-in solution of invenio 3 framework. Not yet linked via machine-actionable API from the metadata GUPRI. Plan to introduce link via "index" metadata of Handle System (http://www.wikidata.org/entity/Q3126718).
	Q17	How are the metadata and data linked? Materials Cloud's FIP presents that a built-in solution of invenio 3 framework is used, and it does not explains how this solution works.
F4	Q18	Which technology is used to make metadata available (or indexed)? The following technologies are used to index metadata: - B2FIND, which is specific for research data. - Google Dataset Search, which offers a great integration with JSON-LD and is easy to setup. - Google, which requires little to no extra effort needed for indexing.
	Q19	How is the metadata available or indexed (*e.g.*, in a search engine, as a static web page, in a database, through an API call)? Not answered considering the FIP.
	Q20	Which technology is used to make your data available (or indexed)? optimade.science, which allows deep searches of atomic structures from optimade-enabled materials databases.
	Q21	How is the data available or indexed (*e.g.*, in a search engine, as a static web page, in a database, through an API call)? Not answered considering the FIP.

Table 4: Questionnaire answers to assess Accessibility considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
A1	Q22	Which is the standardized mechanism or service used to provision the metadata? Open Archives Initiative Protocol for Metadata Harvesting, used for DOI harvesting.
	Q23	Which is the standardized mechanism or service used to provision the data? The following services are used to provision the data: - Optimade API, which allows for interoperable exchange of atomic structure data between databases. - AiiDA REST API, which Allows querying AiiDA provenance graphs stored in AiiDA archives.
A1.1	Q24	Which standardized communication protocols are used to access the metadata? HTTPS
	Q25	Is the protocol used to access the metadata open, free, and universally implementable? Yes, it is.
	Q26	Which standardized communication protocols are used to access the data? HTTPS
	Q27	Is the protocol used to access the data open, free, and universally implementable? Yes, it is.
A1.2	Q28	What are the security mechanisms used for metadata access, such as ones used for authentication and authorization, and access conditions and access levels? Open access [5]
	Q29	What are the security mechanisms used for data access, such as ones used for authentication and authorization, and access conditions and access levels? Open access [5]
	Q30	What security information is provided in the metadata that allows one to access the data manually or through a client application? Not answered considering the FIP.
A2	Q31	Are data and metadata independently stored? Yes, they are.
	Q32	What is the metadata longevity plan? The longevity plan is the Materials Cloud metadata longevity plan. It covers longevity of both metadata and data.
	Q33	What is the data longevity plan, if any? The longevity plan is the Materials Cloud metadata longevity plan. It covers longevity of both metadata and data.

Table 5: Questionnaire answers to assess Interoperability I1 considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
I1	Q34	What is the knowledge representation used for metadata? *E.g.*, Relational, Document, Key Value, Graph, Object, Hierarchical, Network. Not answered considering the FIP.
	Q35	Is the knowledge representation used for metadata formal, accessible, shared, and broadly applicable? Not answered considering the FIP.
	Q36	In what format the knowledge representation used for metadata is provided? *E.g.*, eXtensible Markup Language (XML), Turtle (TTL), JSON, JSON-LD, CSV, BLOB, CLOB. The Materials Cloud FIP does not presents the knowledge representation language. They state the data format they use for data representation, which are: - JSON-LD - JSON, internal representation used by the invenio 3 framework - eXtensible Markup Language, which is used for representation of dublin-core and oai-pmh metadata.
	Q37	Is the format used for knowledge representation of metadata formal, accessible, shared, and broadly applicable? Not answered considering the FIP.
	Q38	What is the knowledge representation used for data? *E.g.*, Relational, Document, Key Value, Graph, Object, Hierarchical, Network. Not answered considering the FIP.
	Q39	Is the knowledge representation used for data formal, accessible, shared, and broadly applicable? Not answered considering the FIP.
	Q40	In what format the knowledge representation used for data is provided? *E.g.*, eXtensible Markup Language (XML), Turtle (TTL), JSON, JSON-LD, CSV, BLOB, CLOB. Not answered considering the FIP.
	Q41	Is the format used for knowledge representation of metadata formal, accessible, shared, and broadly applicable? Not answered considering the FIP.

Table 6: Questionnaire answers to assess Interoperability I2 and I3 considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
I2	Q42	Which structured vocabularies are used to annotate the metadata? The following structured vocabularies are used: schema.org, Dublin Core, and DataCite.
	Q43	Are the used vocabularies for metadata FAIR? Not answered considering the FIP.
	Q44	Which structured vocabularies are used to encode the data? AiiDA Ontology for provenance graphs which is applied only to AiiDA data records.
	Q45	Are the used vocabularies for data FAIR? Not answered considering the FIP.
I3	Q46	Which qualified references the data include to other data? Materials Cloud Archive Record Schema v1.0.0, which is derived from invenio's record schema.
	Q47	Which qualified references the metadata include to other metadata? AiiDA archive format, which includes extensive provenance tracking for full reproducibility. However it is not available for all datasets.

Table 7: Questionnaire answers to assess Reusability considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
R1	Q48	What are the relevant metadata attributes? Not answered considering the FIP.
	Q49	What are the relevant data attributes? Not answered considering the FIP.
	Q50	What is the required accuracy of each attribute, if any? Not answered considering the FIP.
R1.1	Q51	Which usage license is used for your metadata? `https://creativecommons.org/licenses/by-sa/4.0/`, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Results from data mining of the repository should be distributed under an open license as well.
	Q52	Is the metadata usage license clear? Yes, it is.
	Q53	Is the metadata usage license accessible? Yes, it is.
	Q54	Which usage license is used for your data? `https://spdx.org/licenses/Apache-2.0.html` and several others listed in the FIP.
	Q55	Is the data usage license clear? Yes, it is.
	Q56	Is the data usage license accessible? Yes, it is.

Table 8: Questionnaire answers to assess Reusability considering the Materials Cloud's FIP [3].

Principle	Q.ID	Questions
R1.2	Q57	Which metadata schemas do you use for describing provenance of the metadata? Materials Cloud Archive Record Schema v1.0.0, a Built-in solution of invenio 3.
	Q58	Which metadata schemas do you use for describing provenance of the data? AiiDA archive format, which includes extensive provenance tracking for full reproducibility. It is not available for all datasets.
	Q59	What are the attributes used for data provenance? Not answered considering the FIP.
R1.3	Q60	What are the domain-relevant community standards for metadata? Not answered considering the FIP.
	Q61	Do the metadata under assessment meet these domain-relevant community standards? Not answered considering the FIP.
	Q62	What are the domain-relevant community standards for data? Not answered considering the FIP.
	Q63	Do the data under assessment meet these domain-relevant community standards? Not answered considering the FIP.