FAIRification of Multidimensional and Tabular Data by Instantiating a Core Semantic Model with Domain Knowledge: Case of Meteorology

Cassia Trojahn, Mouna Kamel, Amina Annane, Nathalie Aussenac-Gilles, Bao-Long Nguyen, Christophe Baehr

FAIRification of Multidimensional and Tabular Data by Instantiating a Core Semantic Model with Domain Knowledge: Case of Meteorology

Cassia Trojahn[1][0000−0003−2840−005X], Mouna Kamel[1][0000−0002−8739−0747], Amina Annane[2][0000−0002−7223−8478], Nathalie Aussenac-Gilles[1][0000−0003−3653−3223], Bao Long Nguyen[1], and Christophe Baehr[3][0000−0002−1230−893X]

[1] IRIT, CNRS, Université Toulouse 2, France, {prenom.nom}@irit.fr
[2] Geotrend, Toulouse, France amina.anane@geotrend.fr
[3] CNRM, Météo-France, France christophe.baehr@meteo.fr

Abstract. Open data is exposed in several formats, including tabular format. However, the meaning of columns, that can also be seen as dimensions, is not always explicit what makes difficult the reuse of this data for data consumers. This paper presents the FAIRification process of tabular and multidimensional datasets that relies on a (FAIR) core semantic model that is able to represent different kinds of metadata, including the data schema and the internal structure of a dataset. We describe how the instantiation of such a model offers in addition the possibility to describe the semantics of columns using domain ontologies. Once instantiated, this model forms a set of formal metadata that documents the dataset and facilitates understanding by data consumers. This process is then applied to three metereological datasets, for which the degree of improvement of the FAIRness ("I" and "R") has been evaluated.

Keywords: Meteorological data · FAIR principles · Semantic metadata

1 Introduction

Large volumes of Open data, in particular, scientific data shared for an open science, or government and statistical data, are now available on the web. They can be accessed under open licenses from different portals, such as governmental portals for public data (e.g., data.gouv in France[4] or data.gov[5] in the US, European portals like the European Data Portal[6]), portals of public services (e.g., the French National Library[7]), or portals of scientific data (e.g. data-terra.org[8] for Earth Sciences). This data is usually structured in tables, available in various

[4] https://www.data.gouv.fr/fr/
[5] https://www.data.gov/
[6] https://ec.europa.eu/info/statistics/eu-open-data-portal_en
[7] https://data.bnf.fr/
[8] https://www.data-terra.org/

formats, mainly CSV or JSON. Not only the schema of these tables is not always provided or made explicit, but it is also described with properties (in particular the meaning of columns) labelled in a relevant way for domain experts (data producers) but that are not properly understood and reusable by other scientific communities than the one of the authors. For the latter, one of the challenges is to find relevant data among the increasingly large amount of continuously generated data, by moving from the point of view of data producers to the point of view usages. One way to overcome these weaknesses is to guarantee compliance of data to the FAIR principles [22]. These principles correspond to a set of 15 recommendations that aims to facilitate data reuse by humans and machines. The first step towards the fulfilment of FAIR principles is to define precise metadata schemes. Indeed, 12 out of the 15 FAIR principles refer to metadata [22]. To go a step further in improving data FAIRness, several authors have shown that metadata schemes should be based on semantic models (i.e., ontologies) for a richer and more metadata representation [10]. Thanks to their ability to make data types explicit, in a format that can be processed by machines, ontologies are essential to make data FAIR [11]. While most efforts in data FAIRification are limited to specific kinds of metadata, mainly those describing the overall features of datasets and data catalogues, this description is not enough to fully address all FAIR principles [14], in particular for promoting data reuse by other scientific communities.

This paper presents the FAIRification process of tabular and multidimensional datasets using a (FAIR) core semantic model. We describe how the instantiation of such a model additionally provides the ability to describe the semantics of columns using domain ontologies. Once instantiated, this model forms a set of formal metadata (including those describing the data schema and the internal structure of a dataset) that documents the dataset and facilitates understanding by data consumers. We illustrate this process in the meteorological domain. The contributions of the paper are the following: (1) an extension of the work in [21] by describing the late stages of the FAIRification process, i.e. how the core model can be instantiated to generate a domain specific knowledge base (here meteorology) to used as a metadata schema. (2) an extension of the work in [2] by improving the FAIRness of three metereological datasets provided by Météo-France (the official French weather agency) that share the same features. Here, we use a new version of the semantic model [21] that was generalised to accommodate any kind of tabular data together with new notions required to represent dataset collections. (3) an evaluation of the FAIRness degree of different datasets annotated either with existing metadata (most of which are not machine readable) or with semantic metadata using our semantic model, showing how the proposed model improves their interoperability and reusability.

The rest of this paper is organised as follows. Section 2 discusses the main related work. Section 3 shortly presents the used core semantic model and details its instantiation. We expose in Section 4 how this process is performed to describe three datasets in the domain of Meteorology. Section 5 reports the FAIRness

evaluation of these datasets with or without the semantic metadata resulting from the instantiation process. Finally Section 6 concludes the paper.

2 Related work

(FAIR) metadata vocabularies A number of vocabularies has been proposed to represent metadata in general (Dublin core, VoID, Schema.org, DCAT, DCAT-AP)[9], with extensions for accommodating specific kinds of data, such as geo-spatial data (GeoDCAT-AP) or statistical data (StatDCAT-AP). In [17], the authors expose their own way of representing metadata on spatial and temporal data identification, content, distribution and presentation forms. In a different way, [8] extend the existing VoID vocabulary to cover datasets that are not RDF ones. Another group of works and initiatives has addressed the problem of representing domain-specific metadata using domain vocabularies. For instance, in the context of social sciences and humanities, the Data Documentation Initiative[10] (DDI) proposes two XML schemes for metadata, reusing vocabularies like Prov-O[11], DC-terms, Data Cube[12] or CSVW[13]. Targeting tabular data as we do, several proposals have combined the use of RDF Data Cube (qb) with other vocabularies to represent observational data, as in [16] or [24]. Close to our goal, the Semantic Government Vocabulary is dedicated to the annotation of Open Government Data, notably CSV distributions [15]. Thanks to these vocabularies, these authors annotate data in CSV format at different levels of detail and show how this improves the discovery of datasets [15].

FAIR principles and FAIRness evaluation Several frameworks assess the degree of FAIRness of digital objects[14]. The reader can refer to [20] for a recent survey on the topic. In many of them, the evaluation is performed by answering a set of questions – also called metrics or indicators – or by filling a checklist such as the "FAIR Data Maturity Model" [7] or "FAIRshake" [4]. This evaluation can be automated, as proposed by [23, 6], based on web applications that test digital resources against predefined metrics. Recently, in addition to the FAIRness degree of data, the FAIRness of vocabularies and ontologies used as metadata schemas was also evaluated [9, 5]. FOOPS! [18] and O'FAIRe (Ontology FAIRness Evaluator) [1]) are some of the few tools automating this task.

3 FAIRification process

Making data FAIR (FAIRification) can be devided into several steps, such as those of the generic step-by-step FAIRification workflow in [12]: 1) identify the

[9] `https://www.dublincore.org/`, `https://www.w3.org/TR/void/`, `https://www.w3.org/TR/vocab-dcat/`, `https://op.europa.eu/en/web/eu-vocabularies/dcat-ap`

[10] `https://ddialliance.org/learn/what-is-ddi`

[11] `https://www.w3.org/TR/prov-o/`

[12] `https://www.w3.org/TR/vocab-data-cube/`

[13] `http://www.w3.org/ns/csvw#`

[14] most of which are listed here: `https://fairassist.org/`

FAIRification objective, 2) analyze data, 3) analyze metadata, 4) define semantic model for data (4a) and metadata (4b), 5) make data (5a) and metadata (5b) linkable, 6) host FAIR data, and 7) assess FAIR data. As part of a generic methodology, here we use the generic semantic model for metadata proposed in [21]. For an easier understanding of the instantiation process, we briefly present in the following this model and how we instantiate it.

3.1 Core Dataset Metadata Ontology

We defined the Dataset Metadata Ontology Core (*dmo-core*, available at `https://w3id.org/dmo`) to represent data schema and the internal structure of tabular datasets using several FAIR vocabularies. Here we briefly recall the main concepts and the reader can refer to [21] for a detailed description. The notion of *Catalog* is a curated collection of metadata about *Datasets*, which in turn can be described with different types of metadata and may have associated *Distributions*. Distributions may be in a tabular format, where each *Table* is described by its *Schema*. A schema specifies the various *Columns*. Each column has an associated *Measure* or *Dimension*. While the model in [2] focused on the representation of spatio-temporal data (using GeoDCAT-AP and qb4st), *dmo-core* forms a more generic core semantic model for representing any kind of tabular data for any domain by adopting DCAT and qb. Improving the FAIRness of domain datasets requires to instantiate this model, as introduced in the following.

3.2 Model instantiation

The main idea behind our FAIRification process is to associate meaning to the data, in particular the columns of the tabular data. To that extent, domain ontology concepts are associated to that columns. We made the choice of instantiating the model instead of extending it, as there is no need for introducing new concepts or relations, as recommended when reusing the standard vocabularies (CSVW and qb), as detailed in the following.

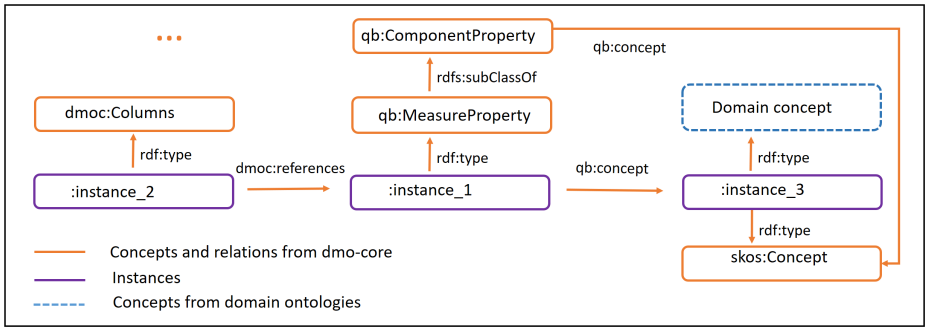

Fig. 1. Instantiation model.

This process can be viewed in two stages: the instantiation of the core model, and the association of domain-specific concepts to the instances. The first stage concerns the description of the tabular dataset as a whole, and reference can be made to [21]. The latter concerns the more specific description of table columns using relevant domain ontologies and is carried out in the following steps:

1. Selecting relevant domain ontologies Several ontology repositories can be queried, such as: Linked Open Vocabularies (LOV) (`https://lov.linkeddata.es/dataset/lov/`), vocab.org (`http://purl.org/vocab/`), ontologi.es (`http://ontologi.es/`), SOCoP+OOR (`https://ontohub.org/socop`), AgroPortal (`https://agroportal.lirmm.fr`), BioPortal (`https://bioportal.bioontology.org/`), OntoHub (`https://ontohub.org/`), COLORE (`http://stl.mie.utoronto.ca/colore/`), OOR (Open Ontology Repository) (`http://www.oor.net/`), ONKI service (`https://onki.fi/`).

2. Choosing appropriate concepts in these ontologies This step requires the intervention of domain experts. In fact, the core model instantiation must be carried out in collaboration between a domain expert and semantic web experts. This collaboration aims to present the domain to the semantic web expert, who is responsible for creating the instantiation files. At this stage, ontology editors such as Protégé[15] can be useful.

3. Associating* dmo-core *to domain concepts Tabular columns (instances of `csvw:Column`) are linked to a dimension, attribute or measure (represented as a `qb:ComponentProperty`). A `qb:ComponentProperty` has a `qb:concept` property whose default range is a `skos:Concept` to which is added a domain concept, as illustrated in Figure 1. More specifically, for each column $\texttt{COL}_i$, the process consists in creating the following instances: (a) $\texttt{INST\_COL}_i$ of `csvw:Column`; (b) an anonymous instance of both `skos:Concept` and the domain concept; (c) $\texttt{COMPONENT\_PROPERTY\_COL}_i$ of `qb:ComponentProperty` (i.e., a dimension, attribute or measure). Then (d) $\texttt{COMPONENT\_PROPERTY\_COL}_i$ is linked to the blank node with the `qb:concept` property. Finally (e) the `dmo-c:references` property links $\texttt{INST\_COL}_i$ to $\texttt{COMPONENT\_PROPERTY\_COL}_i$.

In the example below, to help understanding the meaning of column `t` (:t_col) (Figure 2), from one of the meterological CSV files we evaluate here (Section 4), (:t_col) is associated to the measure :t. This measure is then linked to the `ENVO:ENVO_09200001` concept which represents the `air temperature`.

```
:t rdf:type qb:MeasureProperty;
   qb:concept [
      rdf:type <http://purl.obolibrary.org/obo/ENVO_09200001>,
               <http://www.w3.org/2004/02/skos/core#Concept>
            ] .
:t_col rdf:type csvw:Column;
       dmo-c:references :t.
```

[15] `https://protege.stanford.edu/` (accessed on 28th July 2022.)

4 Metereological datasets FAIRification

Meteorological open data is essential in many applications, including weather forecast, climate change, environmental studies, agriculture, and risk management. Its production is based on mathematical models that assimilate different data from several sources including sensors located on weather stations, satellites and weather radars. While this data has been made available as open data, through different portals, its exploitation is rather limited. Not only the schema of these tables is not always provided or made explicit, but it is also described with properties (in particular the meaning of columns) labelled in a relevant way for meteorology data producers but that are not properly understood and reusable by other scientific communities. For the latter, one of the challenges is to find relevant data among the increasingly large amount of continuously generated data, by moving from the point of view of data producers to the point of view of data usages.

Thus meteorological data is a good experimental ground to test the benefits brought by the addition of semantic metadata based on dmo-core to the dataset reusability by other scientific communities. To this end, we instantiate the *dmo-core* model to describe three collections of tabular datasets (SYNOP, NIVO and SWI) provided by Météo-FranceThese datasets were chosen because, in the context of the Semantics4FAIR[16] project, the biologist partners needed to access and reuse such (understandable) weather data for identifying the meteorological conditions that favor the germination and flowering of ragweed. Currently, on the Météo-France website, these datasets are presented with few metadata in natural language, which prevents dataset search engine crawlers from finding them, and hence minimises the dataset discoverability.

A search for domain ontologies on the above mentioned repositories led us to choose the following ones: SWEET (`http://sweetontology.net/`), ENVO (`http://purl.obolibrary.org/obo/`), QUDT (`http://qudt.org/1.1/vocab/unit`), qb4st (`http://www.w3.org/ns/qb4st/)` and SOSA (`http://www.w3.org/ns/sosa/`). SWEET [19] is a collection of ontologies conceptualizing knowledge for the Earth sciences, a part of which models meteorological parameters such as humidity, wind speed, pressure at sea level or rainfall. ENVO [3] represents environmental entities. It is used in addition to SWEET to better describe environmental processes, for example by offering the possibility to specify the extremes of a temperature (minimum and maximum). QUDT defines the classes, properties, and restrictions for modelling physical quantities, units of measure, and their dimensions in various measurement systems. For our purpose, QUDT allows to specify units of measure for measurements. SOSA [13] is a reference ontology to describe sensors (such as thermometer, barometer, etc.) and their observations (measures), the involved procedures, the studied features of interest, the samples used to do so, and the observed properties, as well as actuators. Finally, qb4st is a qb extension for spatio-temporal components.

[16] `https://www.irit.fr/semantics4fair/`

SYNOP dataset The SYNOP data archive consists of a set of monthly generated files (since January 1996) where each file covers only the observations made in one month. Generated files are freely available online[17]. These files share the same structure (1 table with 59 colums). Figure 2 shows an excerpt of one SYNOP file. DMO-SYNOP corresponds to the instantiation of *dmo-core* to describe the SYNOP dataset. Part of the instantiation is presented below and the whole DMO-SYNOP instantiation is available online[18] with the following prefix `<https://www.irit.fr/recherches/MELODI/ontologies/DMO/dmo-c-synop#>`.

numer_sta	date	pmer	tend	cod_tend	dd	ff	t	td	...
7005	2,02E+13	103180	-80	8	120	1.800000	274.350000	272.750000	...
7015	2,02E+13	103320	0	5	80	4.700000	275.250000	275.150000	...
7020	2,02E+13	102870	-70	8	80	1.300000	280.550000	279.450000	...
7027	2,02E+13	103080	0	0	100	4.200000	275.750000	275.750000	...
7037	2,02E+13	103190	-30	8	130	2.200000	272.250000	272.250000	...
7072	2,02E+13	103320	-20	8	60	1.100000	270.650000	269.550000	...
7110	2,02E+13	102740	10	0	180	0.600000	282.750000	282.650000	...
7117	2,02E+13	102760	-20	8	130	0.500000	281.550000	280.950000	...
7130	2,02E+13	102940	-90	8	110	3.100000	278.350000	278.050000	...
...	...	...	...	...	...	...	...	...	...

Fig. 2. Excerpt of SYNOP data.

Representing metadata of SYNOP dataset. SYNOP dataset is represented by an instance of `dmoc:Dataset`. SYNOP is a collection of monthly files, that, in turn, can be considered as datasets themselves. Using the concepts linked to a dataset in DMO-core, `:SYNOP_dataset` is given the following metadata values: `dct:publisher` is Météo-France; `dct:provenance` is made explicit with the label value "The measurements were provided by the meteo_France stations"; `dct:spatial` points to France in Geonames (<https://www.geonames.org/countries/FR/>); etc. To represent the structure of the dataset, which is shared by all SYNOP files, we use `:SYNOP_dataset_structure`, an instance of `qb:DataStructureDefinition`. `:SYNOP_dataset_structure` is linked to an instance of `qb:ComponentSpecification` for each of the 59 columns, each measure unit (1) and each measuring method (1), i.e. 61 instances in total. For example, the instances `:pmer_Component` and `:month_Component` correspond respectively to "pmer" and "month" columns. Then each of these instance was linked to instances of `qb:MeasureProperty`, `qb:DimensionProperty` or `qb:AttributeProperty` depending on the nature of the component. Finally, these instances are also linked to concepts of domain ontologies (we mainly used SWEET) via the `qb:concept` property.For example, `:pmer_Component` is linked to `sweet:SeaLevelPressure` (<http://sweetontology.net/propPressure/SeaLevelPressure>). The property

[17] `https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32`

[18] `https://www.irit.fr/recherches/MELODI/ontologies/DMO/DMO-core-SYNOP.ttl`

`dmo:requires` makes explicit the dependency between `:SYNOP_dataset` and `:Meteo_Station`, the characteristics (longitude, latitude, etc.) of the weather station generating the measures being stored in the Weather Station file.

Representing metadata of SYNOP dataset (February, 2020). The dataset stored in each file of the SYNOP collection is represented as an instance of `dmo:Slice`. For instance, `:SYNOP_dataset_Feb_20` is an instance of `dmo:Slice` linked to `:SYNOP_dataset` via the property `qb:slice`. `:SYNOP_dataset_Feb_20` is associated with several metadata (`dct:created`, `dct:creator`), including structural metadata via the `qb:structure` property (`:SYNOP_dataset_structure`). Representing the metadata of a `qb:Slice` also requires the definition of dimensions with fixed values, which are specified using the `qb:SliceKey` concept. In our case, the fixed dimensions for a monthly dataset are year and month, with values `month:FEB` and `year:2020`.

Representing metadata of a SYNOP dataset distribution (February 2020). The CSV file itself is represented as a distribution (`dmo:TabularDistribution`) of `:SYNOP_dataset_Feb_20` with identifier `:SYNOP_distribution_Feb_20`. Several metadata associated with this distribution were specified: the format (CSV), the URL from which the CSV file can be downloaded, the kind of license (open license), the description, etc. The distribution schema is represented by `:SYNOP_Schema` (to be reused across distributions), an instance of `csvw:Schema`. It includes all the columns of the CSV file (e.g., `numer_sta` and `pmer`). For each column, we represent its name (`csvw:name`), its label (`csvw:title`), its data type (`csvw:datatype`), etc. The foreign key `:SYNOP_ForeignKey` which connects the column "numer_sta" of the SYNOP data, to the column "ID" of the station data (`:Distribution_Stations_Météo`) is represented by the instance `:SYNOP_Stations_Table_Reference` of `csvw:TableReference`.

NIVO dataset This dataset refers to meteorological observation data from mountain stations operated by partners under agreement with Météo-France for monitoring the snowpack in winter. The generated files, containing data measured since January 1996, are available free of charge online[19], in CSV format. Documentation is available on a PDF file. Each CSV file contains 45 columns (temperature, dew point, snow state, predominant type of surface grains, etc.). DMO-NIVO corresponds to the instantiation of *dmo-core* for describing the NIVO dataset. The instantiation rules are the same as those applied when instantiating DMO-SYMOP. ENVO [3], a knowledge representation of environmental entities, has been used as domain ontology. The whole DMO-NIVO instantiation is available online[20].

[19] `https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=94&id_rubrique=32`

[20] `https://www.irit.fr/recherches/MELODI/ontologies/DMO/DMO-core-NIVO.ttl`

SWI dataset The (uniform) SWI dataset represents the Soil Wetness Index (SWI) calculated by the Safran-Isba-Modcou (SIM) model for measuring complex interactions between meteorological data. This kind of index is used by Météo-France in the reports to commission responsible for the management of natural disasters in France. Generated files are freely available online[21], in CSV format. Each file contains 5 columns: grid cell number, geographic x and y coordinates (Lambert format), date, SWI value. Each monthly value integrates the current month and the two previous months: average of the three of daily SWI values. DMO-SWI corresponds to the instantiation of *dmo-core* for describing the SWI dataset. The instantiation rules and the domain ontology are the same as those applied when instantiating DMO-SYNOP. The whole DMO-SWI instantiation is available online[22].

5 Evaluation

We evaluated the degree of FAIRness of the datasets before and after they are described with *dmo-core*, thanks to the framework *FAIR data maturity model* proposed by the Research Data Alliance (RDA) [7]. This model is based on three components: i) 41 indicators measure the state or level of a digital resource according to a FAIR principle; ii) priorities (*essential*, *important*, *useful*) are associated with the indicators; iii) two evaluation methods: the first assigns each indicator a maturity level between 0 and 4 so that data providers have indications about how to improve the FAIRness degree of their data; the second consists of verifying whether the criterion carried by the indicator is true or false. The indicators were applied first to the original dataset description, and then to the dataset described with metadata (MD) instantiating *dmo-core*. The evaluation was manually carried out and guided by the RDA Excel form.

We first evaluated the original description of the datasets (without semantic metadata). The datasets share the same conditions of access and lack of metadata. Their evaluation confirmed that they were not FAIR: i) level 0 for principles **F**, **A** and **R**, because at least one essential indicator was not satisfied for each of them; ii) level 1 for principle **I**, because no indicator is essential for this principle (Figure 3). The datasets were **re-evaluated** after generating the semantic metadata that describe them. These semantic metadata significantly contribute to improve their FAIRness level, especially for the **I** and **R** principles (Figure 4). In fact, one of the main concerns when proposing the dataset annotation with semantic metadata was to improve their exploitation by non-experts from other scientific communities which would consequently improve their interoperability. Indeed, the proposal meets the main **I** criteria: *metadata and data schemes are expressed in standardised and machine-understandable format, using FAIR-compliant vocabularies; metadata and data refer to other (open) data (here, domain ontologies) and links with these files are made explicit.* Although

[21] `https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=301&id_rubrique=40`

[22] `https://www.irit.fr/recherches/MELODI/ontologies/DMO/DMO-core-SWI.ttl`

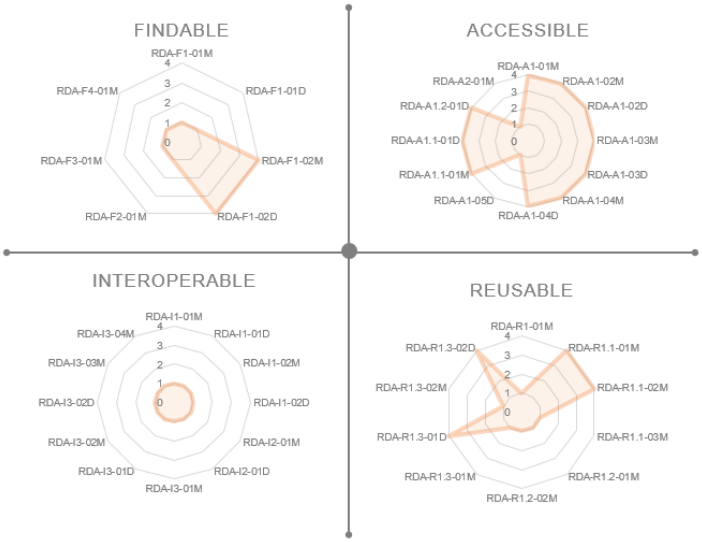

Fig. 3. Dataset FAIRness progress per indicator (without semantic MD)

the re-evaluation of the **F** principle did not show any gain, the model does allow for the representation of "rich" indexing metadata that satisfy **F2** principle. However, higher **F** and **A** degrees would require satisfying essential indicators that are beyond the capabilities of any semantic model e.g., the generation of persistent and unique identifiers (**F1**), persistent metadata (**A2**), publication of metadata on searchable resources (**F4**), which must be managed by the data publisher (Meteo-FR). We also observe that the FAIRness degree is preserved with the generic instantiation model with respect to the results obtained with the specific dmo model in [2].

6 Conclusion and future work

This paper presented the FAIRification process of tabular and multidimensional datasets. It detailed how we defined the metadata of each dataset as instances of the DMO-core ontology and domain-specific ontologies. Three meterological collections of datasets were annotated in that way. The paper finally reported the evaluation of the approach on these meteorological datasets. An evaluation of the FAIRness of the datasets with their semantic metadata proves the relevance of the proposal in the FAIRification process, improving in particular criteria **I** and **R**. Yet we have planned several improvements and additional evaluations in other domains than meteorology. A first goal is to extend *dmo-core* to tabular datasets with other format than CSV, such as XML or JSON (which can be done quite easily by integrating dedicated vocabularies as done for CSVW). In fact, the combination of RDF Data Cube and DCAT is suitable for describing any kind of general metadata. A second one is to write SHACL constraints for the DMO-core ontology and implement a form generated from the SHACL file

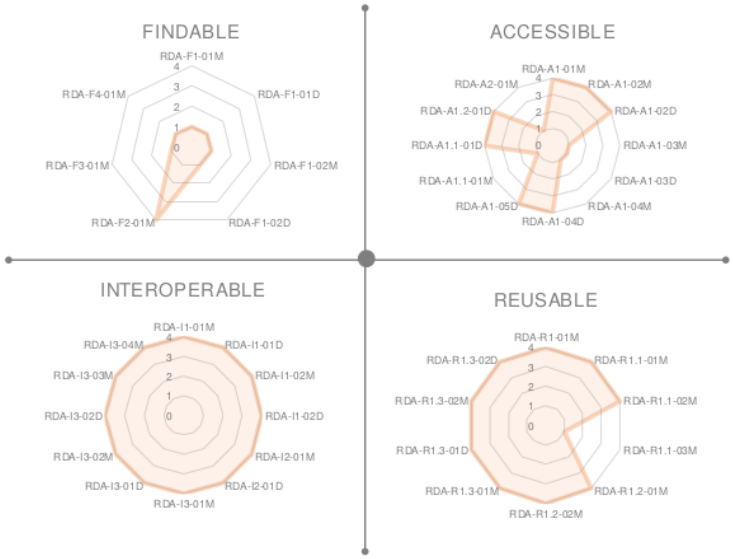

Fig. 4. Dataset FAIRness progress per indicator (with semantic MD)

to make it easier for domain expert to annotate their datasets. a third one could be to test our proposal in other domains (such as health) using other domain ontologies. Finally, we plan to complement our evaluation using other frameworks such as F-uji[23] and Fairshake[24].

References

1. E. Amdouni and C. Jonquet. FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. In *MTSR - 15th International Conference on Metadata and Semantics Research*. Springer, Nov. 2021.
2. A. Annane, M. Kamel, C. Trojahn, N. Aussenac-Gilles, C. Comparot, and C. Baehr. Towards the fairification of meteorological data: A meteorological semantic model. In E. Garoufallou, M.-A. Ovalle-Perandones, and A. Vlachidis, editors, *Metadata and Semantic Research*, pages 81–93, Cham, 2022. Springer.
3. P. L. Buttigieg, N. Morrison, B. Smith, and *et al.* The environment ontology: contextualising biological and biomedical entities. *J. Biomed. Semant.*, 4:43, 2013.
4. D. Clarke and *et al.* Fairshake: Toolkit to evaluate the fairness of research digital resources. *Cell Systems*, 9(5):417–421, 2019.
5. S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, and M.-C. Marinescu. Ten simple rules for making a vocabulary fair. *PLOS Computational Biology*, 17(6):1–15, 06 2021.
6. A. Devaraju, R. Huber, M. Mokrane, P. Herterich, L. Cepinskas, J. de Vries, H. L'Hours, J. Davidson, and A. White. FAIRsFAIR Data Object Assessment Metrics 0.5. Technical report, Research Data Alliance (RDA), Oct. 2020. https://zenodo.org/record/6461229 Accessed 3 May 2022.

[23] `https://www.f-uji.net/`
[24] `https://fairshake.cloud/`

7. FAIR Data Maturity Model Working Group RDA. FAIR Data Maturity Model. Specification and Guidelines, June 2020. https://doi.org/10.15497/rda00050 Accessed 6 May 2022.
8. M. Frosterus, E. Hyvönen, and J. Laitio. Datafinland - A semantic portal for open and linked datasets. In G. Antoniou, M. Grobelnik, and et al., editors, *8th Extended Semantic Web Conference, ESWC, Heraklion, Crete, Greece*, volume 6644 of *LNCS*, pages 243–254. Springer, 2011.
9. D. Garijo and M. Poveda-Villalón. Best practices for implementing FAIR vocabularies and ontologies on the web. *CoRR*, abs/2003.13084, 2020. https://arxiv.org/abs/2003.13084 Accessed May 2022.
10. G. Guizzardi. Ontology, Ontologies and the "I" of FAIR. *Data Intelligence*, 2(1-2):181–191, 2020.
11. A. Jacobsen and *et al.* FAIR principles: Interpretations and implementation considerations. *Data Intelligence*, 2(1-2):10–29, 2020.
12. A. Jacobsen, R. Kaliyaperumal, L. O. B. da Silva Santos, B. Mons, E. Schultes, M. Roos, and M. Thompson. A generic workflow for the data fairification process. *Data Intelligence*, 2(1-2):56–65, 2020.
13. K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, and M. Lefrançois. Sosa: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 56:1–10, 2019.
14. L. Koesten, E. Simperl, T. Blount, E. Kacprzak, and J. Tennison. Everything you always wanted to know about a dataset: Studies in data summarisation. *Int. J. Hum. Comput. Stud.*, 135, 2020.
15. P. Kremen and M. Necaský. Improving discoverability of open government data with rich metadata descriptions using semantic government vocabulary. *J. Web Semant.*, 55:1–20, 2019.
16. L. Lefort, J. Bobruk, A. Haller, K. Taylor, and A. Woolf. A linked sensor data cube for a 100 year homogenised daily temperature dataset. In *Proc. of the 5th Int. Workshop on Semantic Sensor Networks*, volume 904, pages 1–16, 2012.
17. V. Parekh, J. Gwo, and T. W. Finin. Ontology based semantic metadata for geoscience data. In H. R. Arabnia, editor, *Conference on Information and Knowledge Engineering*, pages 485–490, 2004.
18. M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, and Ó. Corcho. Coming to terms with FAIR ontologies. In C. M. Keet and M. Dumontier, editors, *EKAW 2020, Bolzano, Italy*, pages 255–270, 2020.
19. R. Raskin. Development of ontologies for earth system science. In *Geoinformatics: Data to Knowledge*. Geological Society of America, 01 2006.
20. C. Sun, V. Emonet, and M. Dumontier. A comprehensive comparison of automated fairness evaluation tools. In *SWAT4HCLS 2022*, volume 3127, pages 44–53, 2022.
21. C. Trojahn, M. Kamel, A. Annane, N. Aussenac-Gilles, B. L. Nguyen, and C. Baehr. A FAIR Core Semantic Metadata Model for FAIR Multidimensional Tabular Datasets. In *EKAW 2022*, volume (to appear) of *LNCS*. Springer, 2022.
22. M. Wilkinson, M. Dumontier, and *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.
23. M. Wilkinson, M. Dumontier, and *et al.* Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sc. Data*, 6(1):1–12, 2019.
24. N. Yacoubi, C. Faron, F. Michel, F. Gandon, and O. Corby. A Model for Meteorological Knowledge Graphs: Application to Météo-France Observational Data. In *22nd Int. Conf. on Web Engineering, ICWE 2022*, Bari, Italy, July 2022.