# A Taxonomy of Tools and Approaches for FAIRification

Dario Mangione, Leonardo Candela and Donatella Castelli

*Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - Consiglio Nazionale delle Ricerche, Via G. Moruzzi 1, Pisa, 56121, Italy*

## Abstract

The FAIR principles have drawn a lot of attention since their publication in 2016. A broad range of stakeholders is confronting the implementation of these guiding principles in diverse contexts. This paper identifies and discusses the tools and approaches emerging from stakeholders' experiences adopting the FAIR principles in practice. In particular, 225 open access grey literature papers (namely, deliverables, milestones and data management plans) on FAIRification have been scrutinised to infer tools and approaches in use. The wealth of emerging tools (477) has been carefully analysed and organised into a comprehensive map highlighting the significant classes of instruments supporting FAIRification. A critical discussion on this collection of tools and approaches and the FAIRification completes the paper.

## Keywords

FAIR, Survey and overview, Systematic literature review

## 1. Introduction

FAIR principles [1] have been introduced as an essential guide for data producers and publishers in implementing good data management supporting manual and automated deposition, exploration, sharing, and reusing data. In this context, 'data' also refers to all scholarly digital research objects, e.g. algorithms, tools, and workflows leading to a research result, and 'metadata' play a key role. The Digital Library community has a long lasting experience on topics related to FAIRness, e.g. on metadata quality [2]. The initial FAIR principles have been revamped to match better the peculiarity of software [3, 4] and computational workflows [5]. Applying these principles to a significant part of the outputs of the research process is meant to ensure transparency, reproducibility, and reusability.

The need for FAIR data management is widely recognised, and there is a lot of discussion on the identification of actions required to make it the standard practice in science. Some are already in place at a different level of spreading and practice. For example, many funders demand their projects to produce data management plans according to these principles. Tools are emerging that measure the repositories' ability to comply with these principles, and services publicly describe their level of fairness as a measure of quality. These concrete activities are based on

*IRCDL 2022: 18th Italian Research Conference on Digital Libraries, February 24-25, 2022, Padova, Italy*

dario.mangione@isti.cnr.it (D. Mangione); leonardo.candela@isti.cnr.it (L. Candela); donatella.castelli@isti.cnr.it (D. Castelli)

0000-0002-4101-0593 (D. Mangione); 0000-0002-7279-2727 (L. Candela)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

their interpretation of the FAIR objectives and principles to a greater or lesser implementation extent.

Indeed, FAIR principles have been introduced to guide approaches for improving the findability, accessibility, interoperability, and reusability of digital resources [6]. A few years after publishing these principles, due to the growing of meanings associated with them, the need has emerged to clarify what FAIR is not. In the literature and the different presentations on the subject, it is now explained, for example, that "FAIR is not a standard …FAIR is not equal to RDF, Linked Data, or the Semantic Web …FAIR is not just about humans being able to find, access, reformat and finally reuse data …FAIR is not equal to Open …FAIR is not a Life Science hobby" [7]. All in all, communities are called to implement "their own" solution for responding to the FAIR principles. This degree of freedom has led to a proliferation of approaches and technologies around the FAIRification process, i.e. the process of making digital resources FAIR. Although generic workflows governing it have been defined [8], there are no supporting means for communities in identifying suitable technologies and approaches that can be leveraged to respond to community-specific FAIRification needs.

This paper introduces and discusses a taxonomy of tools and approaches exploited in concrete FAIRification activities by analysing a corpus of recently published 225 open access grey literature (namely, data management plans, deliverables, and milestones), leading to the identification of 477 tools. The identified tools are different and support various FAIRification process phases. The produced taxonomy is intended to be a tool itself, supporting communities involved in FAIRification tasks by suggesting common approaches and technologies successfully used by others and gaps be filled by new ones. The paper also critically discusses the current state of the art emerging from such a study, identifying steps still needed to spread the FAIR approach.

The paper is organised as follows. Section 2 describes the methodology underlying this study. Section 3 classifies and describes the approaches and methods for FAIRification exploited by diverse projects and initiatives. Section 4 critically discusses the findings emerging from the investigation. Finally, Section 5 concludes the paper.

## 2. Methodology

A systematic mapping study approach [9] has been exploited to achieve the study's goals. In particular, the study is implemented by a structured process where existing literature relevant to the research topic is identified, categorised, and analysed.

Like any systematic mapping study, the first step consisted in identifying the literature of interest. In particular, the study focused on grey literature. FAIRification is a practical process often documented by Technical Reports, Project Deliverables or Data Management Plans. Moreover, it was decided to leverage OpenAIRE contents to identify the literature of interest because FAIRification is a duty of research projects funded by the European Commission and other funding bodies. These projects are called to make their deliverables available via OpenAIRE. This decision also allows the collection of software artefacts somehow correlated to FAIRification processes.

To identify the literature of interest, we started with the most straightforward query and searched for the terms 'FAIR' and 'FAIRification' and focused on publications having type

‘Report’, ‘Project Deliverable’, and ‘Project Milestone’. The term ‘FAIR’ brings in a lot on “false positive”; thus, we excluded terms like ‘trade’, ‘value’, ‘play’, ‘treatment’, ‘price’, ‘indivisibility’, ‘division’, ‘africa’, ‘tax’, ‘equality’, ‘trial’, ‘admission’, ‘agreement’, ‘payment’. We also restrict the time range of the results to the period 2016-2021 because FAIR principles were officially published in 2016. This process resulted in 487 publications.

To improve the recall and the precision of the search, we contacted OpenAIRE colleagues to be provided with the complete list of keywords and subjects accompanying the selected publications. After revising this list of keywords, we asked for the papers annotated with the chosen keywords. Compared with the previous ones, the results of this “snowballing” allowed us to enrich the initial corpus. The number of identified publications becomes 567 plus 389 software entries explicitly annotated with the selected keywords and subjects.

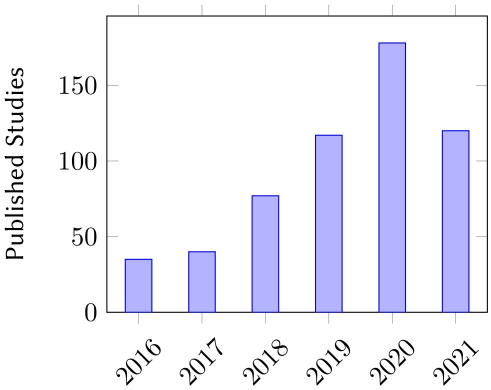

**Figure 1:** Papers in the corpus by publication year

The resulting corpus of grey literature and software is documented by a file accompanying this paper [10].

Not knowing the extent of the references to tools employed for achieving the FAIRness of resources in the identified corpus, whose dates range from 2021 to 2016, we decided to conduct the analysis backwards, starting from the documents published in 2021 until obtaining a significant set of entries that would allow us to build and test a taxonomy of tools.

The sample on which the taxonomy is based consists of 477 unique elements, further reduced to 277 items of immediate interest to the analysis. The difference between the number of items collected and the elements considered valid for the study depended on the willingness to consider tools usually cited in different contexts, as well as the inclusion of elements currently not accessible, not identifiable by their name or acronym, removed or simply with descriptions that did not allow us to assess the relevance of their functions concerning the FAIR principles.

# 3. Analysis

The analysis is based on the manual scrutiny of a refined and final corpus consisting of 225 publications published between 2020 and 2021[1], and 95 software entries, which produced a list of 477 items among tools and services related to the FAIRification process of (meta)data, software and workflows.

The entries were initially organised using the self-assessed information characterising the resources declared by the resource owner, if any. In particular: (*i*) the tool or service category; (*ii*) the reference to the FAIR guiding principle the tool or service responds to by using the corresponding identifier of the FAIR principle; (*iii*) the domain in which the tool or service is used; (*iv*) the FAIRness scope of the tool or service, differentiating among (meta)data [1], software [3, 4], and workflows [5].

Because of the heterogeneity of the results obtained, it became necessary to create an initial classification for normalising the tool/service categories, which became the basis for the development of the taxonomy. As for the normalisation of the declared domains, the Frascati framework is used [11] (see Figure 2 for the actual domain distribution of the tools and services deemed valid for creating the taxonomy).

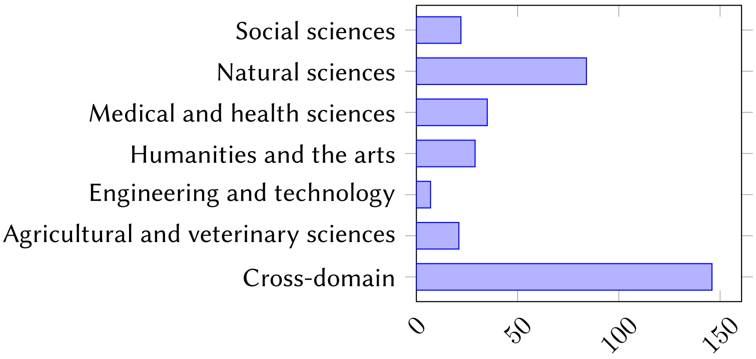

**Figure 2:** Domain distribution of tools and services

## 3.1. Initial FAIR principle-based categories

Each FAIR principle was analysed from the point of view of the class of tools and services required for its implementation (see Table 1 below). The discussions and clarifications given in [6] were taken into account.

Eight of the fifteen FAIR principles (namely, F1, F2, F3, F4, I1, R1.1, R1.2, and R1.3) led us to create five candidate classes: (*i*) *GUPRI creation and management service*; (*ii*) *Metadata helper*; (*iii*) *Indexing and discovery service*; (*iv*) *Licence helper*; (*v*) *Converter*. To the first five FAIR principles-driven classes, we added (*vi*) *Assessment tool* to include the tools and services used to assess a resource's overall FAIRness and, consequently, interesting all FAIR recommendations

[1] We observed that publications preceding it were mainly referring to either tools already referred by publications in the corpus or tools no longer existing or superseded by new ones.

**Table 1**
FAIR principles and corresponding initial categories of tools and services

| FAIR Principle | Tool/service category |
|---|---|
| F1: (Meta) data are assigned globally unique and persistent identifiers | GUPRI helper |
| F2: Data are described with rich metadata | Metadata helper |
| F3: Metadata clearly and explicitly include the identifier of the data they describe | Metadata helper |
| F4: (Meta)data are registered or indexed in a searchable resource | Indexing & discovery service |
| A1: (Meta)data are retrievable by their identifier using a standardised communication protocol | |
| A1.1: The protocol is open, free and universally implementable | |
| A1.2: The protocol allows for an authentication and authorisation where necessary | |
| 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 | Metadata helper, Converter |
| I2: (Meta)data use vocabularies that follow the FAIR principles | |
| I3: (Meta)data include qualified references to other (meta)data | Metadata helper |
| R1: (Meta)data are richly described with a plurality of accurate and relevant attributes | |
| R1.1: (Meta)data are released with a clear and accessible data usage licence | Licence helper |
| R1.2: (Meta)data are associated with detailed provenance | Metadata helper |
| R1.3: (Meta)data meet domain-relevant community standards | Metadata helper, Converter |

**Table 2**
Initial main categories of tool and services

| Tool/service main category | FAIR Principle reference |
|---|---|
| GUPRI helper | F1 |
| Metadata helper | F2, F3 |
| Indexing and discovery service | F4 |
| Metadata helper | |
| Converter | I1, I3 |
| Licence helper | R1.1 |
| Metadata helper | R1.2 |
| Converter | R1.3 |
| Assessment tool | F A I R |

(see Table 2). The following sections present the resulting taxonomy and describe the identified classes.

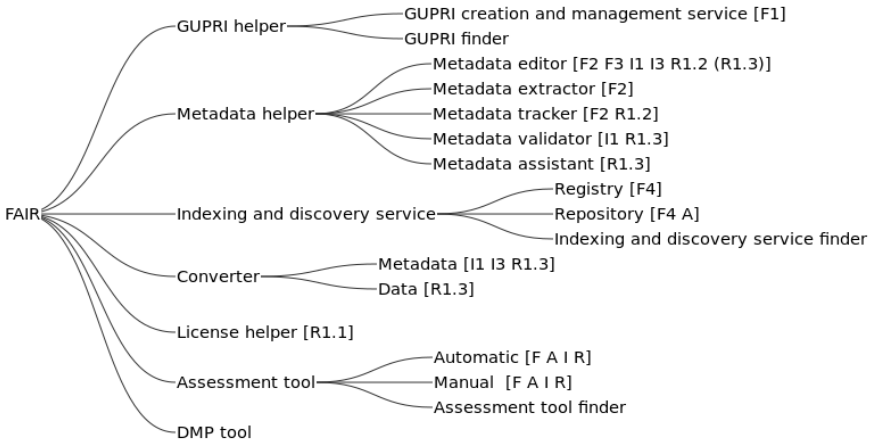

**Figure 3:** Taxonomy of FAIR tools

## 3.2. Taxonomy of FAIR tools

By analysing and integrating the 477 tool and service entries obtained through the scrutiny of the grey literature, we have created a taxonomy of FAIR tools structured in seven main classes: (*i*) *GUPRI helper*; (*ii*) *Metadata helper*; (*iii*) *Indexing and discovery service*; (*iv*) *Converter*; (*v*) *Licence helper*; (*vi*) *Assessment tool*; (*vii*) *DMP tool*.

The seventh category, DMP tool, was added at a later stage (i.e. during the analysis of the collected items) to accommodate the tools that allow the creation of data management plans. Although debatable, this decision was taken because, despite these tools not directly impact on the FAIRness of a resource, they nevertheless provide a coherent reference framework for its management and consequently the possibility of its adherence to the FAIR guidelines.

The seven main classes are generally disjoined except for Metadata helper and Converter, since there are tools that fall under the first category that also include a converter function, i.e. a metadata helper can also be a converter, although this functionality is not its focal role.

Figure 3 also associates the categories of tools with the FAIR principles they contribute to.

### 3.2.1. GUPRI helper

Globally unique, persistent and machine-resolvable identifiers (GUPRI) are the primary elements to be set for both data and metadata being suggested by the F1 principle. Every FAIRness activity is called to use tools and approaches helping to generate unambiguous identifiers that will continue to work even when the “asset” target of the FAIRification activity is going to disappear, thus becoming no longer available. Because of this role, services and technologies for GUPRI should guarantee the long term availability of the identifiers assigned to the FAIRified assets. The need to have identifiers with specific characteristics was discussed in previous works [12, 13, 14]. In particular, these earlier works highlighted how web-based identifiers were with us well before the FAIR advent and that their core role and efficacy is guaranteed only under some settings (namely, the sustainability and governance of the specific system).

This category consists of 33 instances divided between (*i*) *GUPRI creation and management service* (32 entries); and (*ii*) *GUPRI finder* (1 entry). The first category includes the providers offering GUPRI registration services, with Handle system implementations (e.g. DOI, ePIC, EUDAT B2HANDLE) and Open Researcher and Contributor ID (ORCID) being the most cited. The latter consist of the services that offer a registry providing indexing and search capabilities for finding a GUPRI-related provider, with PID Services Registry [15] being its only instance.

The Handle System [16] is a proprietary registry of the Corporation for National Research Initiatives (CNRI) administered by the DONA Foundation, providing a resolution system upon which single Handle.Net software instances are built. The DOI System [17], developed and maintained by the International DOI Foundation, is the most cited implementation of the Handle system and among GUPRIs in the corpus. It ensures the persistence of the handles through a federation of registration agencies, among which the most cited in the documents are Crossref [18] and DataCite [19], mainly via Zenodo.

### 3.2.2. Metadata helper

While, ontologically speaking, the existence of a GUPRI is enough to assert that something unique exists and being resolvable it also allows to locate it, metadata enables to find and identify the associated resource by filtering among similar resources through the accumulation of characteristics.

Metadata helpers allow defining the metadata accompanying a resource or altering already existing ones, both manually or automatically, whether they are embedded or stored as a separate file. They also allow the identification of possible metadata directly from the resource to enrich its description, to suggest the appropriate value for a metadata field or check the conformance of the existing metadata against a standard.

Based on the 65 tools and services we found in the grey literature and on the functions above, we distinguish metadata helpers among five subclasses, namely (*i*) *Metadata editor* (38 entries); (*ii*) *Metadata extractor* (6 entries); (*iii*) *Metadata tracker* (13 entries); (*iv*) *Metadata validator* (7 entries); and (*v*) *Metadata assistant* (1 entry).

Following the FAIR principles order, at the most basic level the metadata editors allow uncontrolled user input, enabling F2, F3 and R1.2, but they can include functions to enable I1, I3 and/or R1.3 by referring to knowledge representation languages and community standards and by validating and/or suggesting the input. They can also include functions to convert between knowledge representation languages, metadata schemas or file formats, thus also pertaining to the converter category. OpenRefine [20], for example, can be seen as a metadata editor, allowing to simply alter a description template or, by using an RDF (Resource Description Framework) extension, as a metadata converter since it enables its user to export the structured data to RDF.

The metadata editor category is quite heterogeneous since it encompasses tools from spreadsheets template generators (like the SIOS Excel template generator [21] for creating Darwin Core compliant descriptions) to mapping tools (like the CIDOC CRM-oriented FORTH-ICS Mapping Memory Manager (3M) [22], which uses the 3M Editor for assisting in the creation of the mappings, suggesting and validating the user input). Moreover, it is not uncommon for metadata editors to offer an integrated validation tool for checking the compliance of the output to a metadata schema, and generally for metadata helpers to share functions. For these reasons, the subclasses of the metadata helpers are usually not disjoint; their instances are classified on a prime function basis.

Metadata extractor encompasses the tools that enable parsing, identification and extraction of metadata, ranging from general-purpose extractors like Apache Tika [23] to metadata specific ones, like Nanopub JupyterLab Extension [24], which helps to extract Nanopublications from a python notebook.

The metadata tracker category is strictly linked with the R1.2 principle since it consists of tools that allow metadata collection in concomitance with the data workflow, including information on data acquisition, generation and processing. While the functionalities of these tools may vary, all of them share the capability of enabling an environment that fosters traceability and reproducibility at least at some point during the data lifecycle. The category includes nbcomet [25], which is a Jupyter notebook extension that regularly saves snapshots and logs every action performed in the notebook, but also workflow managers like Taverna [26] and data management systems like openBIS (open Biology Information System) [27], that automates (meta)data ingestion while providing data provenance tracking and which is designed to be integrated with workflow managers.

Metadata validators check the conformance of a resource to a metadata standard, contributing to the enablement of the I1 and R1.3 principles. This category encompasses single-standard oriented validators, for instance (*a*) the METS Validator [28] provided by the Finnish Digital Preservation Service for Research Data, that checks a METS file against the national digital preservation specification, or (*b*) the CF (Climate and Forecasts) Checker [29], that validates a netCDF file against the CF metadata conventions, and multi-format checkers like OCTOPUS [30], which validates against the SeaDataNet ODV, netCDF and MedAtlas standards. The latter is also an example of a multifunctional tool, combining a validator, a converter from and to SeaDataNet standards, an editor for splitting a SeaDataNet file and an extractor.

The last subclass of the metadata helper category is metadata assistant. This category includes tools that help insert metadata based on controlled vocabularies, contributing to the standardisation of the descriptions produced and, consequently, the fulfilment of the R1.3 principle. While some metadata editors include an auto-completion feature, the only instance found in the examined literature that is dedicated to this task is CEDAR OnDemand [31], a browser extension that helps standardise repository descriptions by recognising the input fields and suggesting the relevant terms from NCBO (National Center for Biomedical Ontology) BioPortal vocabularies.

### 3.2.3. Indexing and discovery service

The Indexing and Discovery service category is by far the most populated category in the taxonomy, consisting of 123 entries distributed in three subclasses; (*i*) *Registry* (54 entries); (*ii*) *Repository* (67 entries); and (*iii*) *Indexing and discovery service finder* (4 entries). It includes all services and tools used for indexing metadata, for discovering the related resources, ultimately enabling the F4 principle and covering the whole accessibility, as well as those that allow the indexing and discovery of the services themselves (e.g. re3data [32] enabling data repositories discovery). These tools are characterised by the modalities of access to the resources they provide and the type of resources managed.

By modality of access, it is possible to distinguish between repositories (e.g. Zenodo [33] or FigShare [34]), which store, index and allow the discovery of resources, and registries (e.g. BARTOC.org [35] or bio.tools [36]), which index metadata and allow the discovery of resources from different repositories. Registries do not store resources on their own rather, they refer to the specific repositories for access and ultimately constitute a catalogue consisting of the metadata describing the resources.

The distinction between repositories and registries become blurred when referring to semantic artefacts [37]. In this case, the metadata registry and metadata repository categories tend to overlap since metadata schemas can be stored in a database by registering the element sets and their constituting elements. Still, it was adopted as a distinguishing characteristic of a semantic artefact repository the access to single elements of a semantic artefact. Based on this distinction, the above-mentioned BARTOC.org (Basic Register of Thesauri, Ontologies & Classifications) is considered a registry for semantic artefacts. At the same time, the NERC Vocabulary Server (NVS) [38] is categorised as a semantic artefact repository.

The analysis showed that there are four types of resources managed by indexing and discovery services, namely (*i*) *data*, (*ii*) *semantic artefacts*, defined as "machine-actionable and -readable formalisation of a conceptualisation, enabling sharing and reuse by humans and machines" [37] (e.g. thesauri, ontologies), (*iii*) *software*, and (*iv*) *workflows*. It is possible to specialise the repository and registry classes further to highlight the types of resources they focus on, although catch-all ones exist. WorkflowHub [39] is an example of a registry dedicated to computational workflows, while GitHub [40] is a representative case of a software repository.

Zenodo is, without any doubt, the most cited repository solution in the corpus. It is an example of a catch-all repository accepting all of the mentioned types of resources. By assigning a DOI to every registered resource, it is also used as a FAIRifying solution in combination with other services like GitHub, which does not reserve a GUPRI for the deposited code, or ARGOS [41], that uses Zenodo for publishing DMPs.

### 3.2.4. Converter

The converter category includes the tools and services that convert data or metadata between models and formats, enabling the transition to community-adopted standards and the combination of resources across different domains and organisations. This class of tools is linked to the I1 and R1.3 principles.

It consists of 39 tools that, following the distinction between (meta)data formats and data models, are arranged in two subclasses: (*i*) *data converter* (26 entries), transforming data between file formats, enabling R1.3, and (*ii*) *metadata converter* (13 entries), that enables I1 and/or R1.3 principles by allowing the transformation to and between knowledge representation languages.

Elements of the first subclass are ImageMagick, which among its functions allows converting images from and to a multitude of formats, and Tabula [42], which transforms data rows in a textual PDF file into a CSV.

Examples of the second subclass are OpenRefine and the similar excel2rdf [43]. The latter tool allows the conversion of an Excel-based vocabulary to a SKOS RDF one, and between metadata formats, like CMD2DC [44], which transforms the CMD resource descriptions used by the CLARIN's Component Metadata Infrastructure (CMDI) to Dublin Core.

### 3.2.5. Licence helper

This category, consisting of 4 entries, includes the tools that help choose a licence for a resource by answering a questionnaire, , facilitating the R1.1 principle. The questionnaire can be more or less detailed, depending on the number of licences taken into consideration, including the type of resource, actual ownership, identification, use and distribution requirements, and providing in some instances the possibility to obtain a machine-readable version of it. For example, the Creative Commons' License Chooser [45] lets users select the most appropriate licence among the six Creative Commons licence types. In contrast, the EUDAT B2SHARE license selector [46] allows users to choose among twenty-two different licences, also distinguishing between software and data-oriented licences. Both produce a machine-readable version of the chosen licence, the first in XMP format, the latter in JSON.

### 3.2.6. Assessment tool

The assessment of the FAIRness of a resource is not strictly a FAIRifying function per se. It does not contribute to implementing any of the fifteen guiding lines. However, assessment tools allow the uptake of the FAIR principles by validating the overall conformity of a resource to a set of criteria or metrics.

Based on the clarity, granularity and measurability of the implemented metrics, and ultimately on their machine-actionability, the evaluation process can be automated or manual. In both cases, it consists in following a questionnaire-like approach. For automated tools, the objectivity of the evaluation criteria allows effective feedback on the resource FAIRness. For manual tools, being a self-assessment, it is more a matter of encouraging the data curators' awareness of the FAIR principles.

Tools and services in this category are consequently cross-principles and specialised into (*i*) *Automated assessment tool* (4 entries), when the machine-accessible metadata of the resource is automatically compared with predefined metrics following the submission of the GUPRIs identifying the resources; (*ii*) *Manual assessment tool* (7 entries), if the FAIRness score of a resource is based on manually filling in a questionnaire; and ( (*iii*) *Assessment tool finder* (1 entry), encompassing the services dedicated to the discovery of assessment tools.

F-UJI and FAIR-Aware [47] respectively exemplify well the first two categories. Developed by FAIRsFAIR and based on fifteen core metrics that, being aligned with the FAIR principles and the CoreTrustSeal requirements, systematically measure the extent to which research data objects are FAIR. F-UJI is an automated assessment tool that evaluates the FAIRness of datasets against the FAIRsFAIR Data Object Assessment Metrics, according to the aggregated resource metadata, through GUPRI creation and management services and repository indexing and discovery services. FAIR-Aware helps researchers understand how to increase the FAIRness of a dataset before depositing it in a repository by a ten-step questionnaire.

Finally, the assessment tool finder category encompasses the indexing and discovery services dedicated to the FAIR assessment tool, with FAIRassist [48] (developed as a component of FAIRsharing [49]) as the only FAIR assessment tool registry found in the examined corpus.

### 3.2.7. DMP tool

As previously mentioned, the inclusion in the presented taxonomy of a DMP tool category may be questionable since its services do not act directly on the resources. However, we decided to incorporate it because, by creating a reference for the resource lifecycle management, establishing how the resources have to be stored, curated, shared and preserved, these tools affect the FAIRness of a resource, particularly its reusability [50, 51].

This class includes eight services that noticeably share the same main function of providing step-by-step guidance in creating a data management plan by filling in annotated forms. Still, they may vary appreciably in additional functions, for instance by supporting machine actionability, allowing to simultaneously collaborate to the realisation of the DMP, offering customisable templates or providing a platform to share them. For example, the ARGOS tool developed by OpenAIRE and based on OpenDMP allows the collaborative creation of a machine-actionable DMP by supporting its versioning, the assignment of a licence and a DOI (through Zenodo) and its publishing.

They can also vary in domain coverage like the ARIADNEplus DMP Researcher Template for Archaeological Datasets that is manifestly domain-specific.

## 4. Discussion

The analysis conducted in this study highlights how FAIRness approaches are envisaged and implemented in recent projects and initiatives. Although it is not possible to claim that the conducted research is exhaustive of the topic of FAIRness, the implemented methodology guarantees high coverage concerning FAIRness initiatives and related tools.

The problem addressed by the study is intrinsically complex because interpretations of FAIR principles and solutions often depend on settings characterising the application context.

A deeper analysis would need to complement the knowledge obtained from the selected documents with details collected "in the field" on how the various communities have practically decided to organise themselves to implement the FAIR principles.

The "interpretability" of the FAIR principles [6] introduces vagueness making it challenging to compare diverse experiences and uses of the tools.

The continuous evolution represents another element to be taken into account. Communities' plans and approaches might be rethought to meet better FAIR principles and expectations resulting from early attempts to make data (more generally, resources) FAIR and to exploit released data. As a consequence of these evolutions, new tools may emerge.

The nature and heterogeneity of the examined material made it possible to observe trends that would otherwise be difficult to appreciate.

There is a tendency to include concepts that are not strictly pertinent to the FAIR principles, although related. For example, accessibility often overlaps with open accessibility, a concept related to open access that does not find a direct acknowledgement in the FAIRness since FAIR data does not mean open data [7]. Open accessibility is linked more to the licences associated with the resources, hence to reusability, than to accessibility itself, which is defined without mentioning open access. Similarly, it is likely to see the concept of reusability intuitively associated with preservation, pursued by depositing the resources into a repository, even if the latter is not mentioned in the definition of reuse given by the principles.

Privileging certain elements of FAIRness over others is another clear trend. For instance, in the case of reuse point R1.1, and therefore the aspect related to the licence, it is mentioned while neglecting the others.

It is also possible to see unexpected tools mentioned in a FAIRness context, e.g. tools linked to collaboration or dissemination, including content management systems, wikis, infrastructures and journals.

By not specifying implementations, the fifteen points in which the guidelines are articulated remain substantially open to interpretation [52] from which this study is no exception. The categories in which the taxonomy is structured result from an analysis of the FAIR principles, which led us to create an equivalence relation between the tool functions and the fifteen guidelines.

While the F2 principle states that the concept of rich metadata is defined in R1, thus creating a direct link between findability and reusability, our interpretation of the F2 principle limits the "rich metadata" only to those that allow to find a resource and to distinguish it from similar ones. In fact, it is arguable that metadata enabling findability should be a subset of those contributing to the reusability of a resource and as such the minimum requirements for its description, that should be defined on a community basis. This view is based on a vision of the FAIR principles as incremental tasks mainly build upon the metadata associated with the resources. Therefore, the F2-related tools and services are not always associated also with reusability.

Similarly, the distinction between tools enabling I1 and those that foster R1.3 is founded on the difference between knowledge representation languages, including their serialisations, and metadata formats. Following this rationale, a tool enabling machine-actionability, without specifying the metadata standard used, is considered just I1-related.

Moreover, some principles are just cited as a general reference, as in the case of accessibility, or are not associated with any tool, such as I2. That is because accessibility depends directly on F1 and on the protocols and policies implemented by registries and repositories rather than a category of tools, while I2 introduces a recursive element among the principles without creating the need for additional categories.

An interesting point highlighted by this study, albeit not unexpected given the number of different implementation contexts and the inherent complexity of the matter, is the lack of unique solutions for ensuring the FAIRness of resources. Even the most integrated solutions found in the corpus (for example, Fairdata.fi [53], which is a nationwide solution promoted by the Finnish Ministry of Education and Culture providing services for storing, describing, indexing, searching and publishing research data) can not cover all of the specific needs arising from the different resource types and scientific workflows.

The need to support the sharing of FAIRness implementation experiences and solutions is actual, and concrete approaches are needed, as demonstrated in [52, 54]. The FAIR Convergence Matrix [52] is a collaborative online resource aiming at creating machine-actionable descriptions of FAIR implementation choices made by different domain communities. The concept of FAIR Implementation Profile [54] aims at capturing by a FAIR object itself the comprehensive set of implementation choices made at the discretion of individual communities of practice. These tools are envisaged to be used to track the evolving landscape of FAIR implementations and inform about them.

## 5. Conclusion and Prospects

Several initiatives are promoting the FAIRification of science assets to improve their findability, accessibility, interoperability and reusability for both humans and machines. Communities responsible for these assets are implementing diverse strategies and approaches to respond to the FAIR guiding principles.

This paper reported the result of a systematic collection and analysis of the tools developed and exploited by scientific communities in their FAIRification activities. A taxonomy organising the various tools into seven major classes is introduced and discussed. This taxonomy is a helpful instrument for understanding the state of the art regarding FAIRification activities, supporting communities of practice confronting with FAIRification activities and helping to develop innovative solutions and strategies improving and filling existing gaps in supporting FAIRification processes.

To further develop the taxonomy, it is planned to systematically assess its efficacy in concrete settings by establishing a dialogue with communities of practice and initiatives engaged in FAIRness activities and systematically analysing the FAIRness approaches documented by FAIR Implementation Profiles [54].

## Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under Blue Cloud project (grant agreement No. 862409), DESIRA project (grant agreement No. 818194), EOSC-Pillar project (grant agreement No. 857650), and SoBigData-PlusPlus (grant agreement No. 871042).

**Author Contributions** According to CRediT taxonomy, authors contributed as follows: DM performed Methodology, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; LC performed Conceptualization, Methodology, Writing - Review & Editing, Supervision, and Funding acquisition; DC performed Conceptualization, Writing - Review & Editing, Supervision, and Funding acquisition.

## References

[1] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding

principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.

[2] A. Tani, L. Candela, D. Castelli, Dealing with metadata quality: The legacy of digital library efforts, Inf. Process. Manag. 49 (2013) 1194–1205. URL: https://doi.org/10.1016/j.ipm.2013.05.003. doi:10.1016/j.ipm.2013.05.003.

[3] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. van de Sandt, J. Ison, P. A. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J. L. Gelpi, N. Chue Hong, C. Goble, S. Capella-Gutierrez, Towards FAIR principles for research software, Data Science 3 (2020) 37–59. doi:10.3233/DS-190026.

[4] D. S. Katz, M. Gruenpeter, T. Honeyman, Taking a fresh look at FAIR for research software, Patterns 2 (2021) 100222. doi:10.1016/j.patter.2021.100222.

[5] C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, D. Schober, FAIR Computational Workflows, Data Intelligence 2 (2020) 108–121. doi:10.1162/dint_a_00033.

[6] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, C. Goble, G. Guizzardi, K. K. Hansen, A. Hasnain, K. Hettne, J. Heringa, R. W. Hooft, M. Imming, K. G. Jeffery, R. Kaliyaperumal, M. G. Kersloot, C. R. Kirkpatrick, T. Kuhn, I. Labastida, B. Magagna, P. McQuilton, N. Meyers, A. Montesanti, M. van Reisen, P. Rocca-Serra, R. Pergl, S.-A. Sansone, L. O. B. da Silva Santos, J. Schneider, G. Strawn, M. Thompson, A. Waagmeester, T. Weigel, M. D. Wilkinson, E. L. Willighagen, P. Wittenburg, M. Roos, B. Mons, E. Schultes, FAIR Principles: Interpretations and Implementation Considerations, Data Intelligence 2 (2020) 10–29. doi:10.1162/dint_r_00024.

[7] B. Mons, C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, M. D. Wilkinson, Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the european open science cloud, Information Services & Use 37 (2017) 49–56. doi:10.3233/ISU-170824.

[8] A. Jacobsen, R. Kaliyaperumal, L. O. B. da Silva Santos, B. Mons, E. Schultes, M. Roos, M. Thompson, A Generic Workflow for the Data FAIRification Process, Data Intelligence 2 (2020) 56–65. doi:10.1162/dint_a_00028.

[9] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, EASE'08, BCS Learning & Development Ltd., Swindon, GBR, 2008, pp. 68–77.

[10] D. Mangione, L. Candela, D. Castelli, A taxonomy of tools and approaches for fairification, 2022. doi:10.5281/zenodo.6037508.

[11] OECD, Frascati Manual 2015, Guidelines for Collecting and Reporting Data on Research and Experimental Development,, OECD Publishing, 2015. URL: https://www.oecd-ilibrary.org/content/publication/9789264239012-en. doi:10.1787/9789264239012-en.

[12] J. A. McMurry, N. Juty, N. Blomberg, T. Burdett, T. Conlin, N. Conte, M. Courtot, J. Deck, M. Dumontier, D. K. Fellows, A. Gonzalez-Beltran, P. Gormanns, J. Grethe, J. Hastings, J.-K. Hériché, H. Hermjakob, J. C. Ison, R. C. Jimenez, S. Jupp, J. Kunze, C. Laibe, N. Le Novère, J. Malone, M. J. Martin, J. R. McEntyre, C. Morris, J. Muilu, W. Müller, P. Rocca-Serra, S.-A. Sansone, M. Sariyar, J. L. Snoep, S. Soiland-Reyes, N. J. Stanford, N. Swainston,

N. Washington, A. R. Williams, S. M. Wimalaratne, L. M. Winfree, K. Wolstencroft, C. Goble, C. J. Mungall, M. A. Haendel, H. Parkinson, Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data, PLOS Biology 15 (2017) 1–18. URL: https://doi.org/10.1371/journal.pbio.2001414. doi:10.1371/journal.pbio.2001414.

[13] N. Juty, S. M. Wimalaratne, S. Soiland-Reyes, J. Kunze, C. A. Goble, T. Clark, Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR, Data Intelligence 2 (2020) 30–39. URL: https://doi.org/10.1162/dint_a_00025. doi:10.1162/dint_a_00025.

[14] J. Klump, R. Huber, 20 years of persistent identifiers – which systems are here to stay?, Data Science Journal 16 (2017). doi:10.5334/dsj-2017-009.

[15] DataCite, 2022, PID services registry, URL: pidservices.org.

[16] Corporation for National Research Initiatives, 2022, Handle.net registry, URL: handle.net.

[17] International DOI Foundation, 2022, The doi system.

[18] G. Hendricks, D. Tkaczyk, J. Lin, P. Feeney, Crossref: The sustainable source of community-owned scholarly metadata, Quantitative Science Studies 1 (2020) 414–427. doi:10.1162/qss_a_00022.

[19] J. Brase, Datacite - a global registration agency for research data, in: 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, 2009, pp. 257–261. doi:10.1109/COINFO.2009.66.

[20] R. Verborgh, M. De Wilde, Using OpenRefine, Packt, 2013.

[21] Svalbard Integrated Arctic Earth Observing System, 2022, Nansen legacy excel template generator, URL: https://sios-svalbard.org/cgi-bin/darwinsheet/index.cgi.

[22] Y. Marketakis, N. Minadakis, H. Kondylakis, K. Konsolaki, G. Samaritakis, M. Theodoridou, G. Flouris, M. Doerr, X3ML mapping framework for information integration in cultural heritage and beyond, International Journal on Digital Libraries 18 (2017) 301–319. URL: https://doi.org/10.1007/s00799-016-0179-1. doi:10.1007/s00799-016-0179-1.

[23] C. A. Mattmann, J. L. Zitting, Tika in Action, Manning, 2011.

[24] R. Richardson, 2021, Nanopubjl, URL: https://github.com/fair-workflows/NanopubJL/tree/v0.3.0.

[25] A. Rule, 2022, nbcomet, URL: https://github.com/activityhistory/nbcomet.

[26] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat, P. Li, Taverna: a tool for the composition and enactment of bioinformatics workflows, Bioinformatics 20 (2004) 3045–3054. URL: https://doi.org/10.1093/bioinformatics/bth361. doi:10.1093/bioinformatics/bth361.

[27] A. Bauch, I. Adamczyk, P. Buczek, F.-J. Elmer, K. Enimanev, P. Glyzewski, M. Kohler, T. Pylak, A. Quandt, C. Ramakrishnan, C. Beisel, L. Malmström, R. Aebersold, B. Rinn, openBIS: a flexible framework for managing and analyzing complex data in biology research, BMC Bioinformatics 12 (2011) 468. URL: https://doi.org/10.1186/1471-2105-12-468. doi:10.1186/1471-2105-12-468.

[28] National digital preservation services, 2022, METS Validation Tool, URL: https://www.digitalpreservation.fi/en/mets-validator.

[29] R. Hatcher, 2022, Cf checker, URL: https://github.com/cedadev/cf-checker.

[30] SeaDataNet, 2022, OCTOPUS, URL: https://www.seadatanet.org/Software/OCTOPUS.

[31] S. A. C. Bukhari, M. Martínez-Romero, M. J. O'Connor, A. L. Egyedi, D. Willrett, J. Graybeal,

M. A. Musen, K.-H. Cheung, S. H. Kleinstein, CEDAR OnDemand: a browser extension to generate ontology-based scientific metadata, BMC Bioinformatics 19 (2018) 268. URL: https://doi.org/10.1186/s12859-018-2247-6. doi:10.1186/s12859-018-2247-6.

[32] H. Pampel, P. Vierkant, F. Scholze, R. Bertelmann, M. Kindling, J. Klump, H.-J. Goebelbecker, J. Gundlach, P. Schirmbacher, U. Dierolf, Making research data repositories visible: The re3data.org registry, PLOS ONE 8 (2013) 1–10. URL: https://doi.org/10.1371/journal.pone.0078080. doi:10.1371/journal.pone.0078080.

[33] European Organization For Nuclear Research, OpenAIRE, Zenodo, 2013. URL: https://www.zenodo.org/. doi:10.25495/7GXK-RD71.

[34] M. Hahnel, 2022, figshare, URL: figshare.com.

[35] J. Waeber, A. Ledl, A semantic web skos vocabulary service for open knowledge organization systems, in: E. Garoufallou, F. Sartori, R. Siatri, M. Zervas (Eds.), Metadata and Semantic Research, Springer International Publishing, Cham, 2019, pp. 3–12. doi:10.1007/978-3-030-14401-2_1.

[36] J. Ison, K. Rapacki, H. Ménager, M. Kalaš, E. Rydza, P. Chmura, C. Anthon, N. Beard, K. Berka, D. Bolser, T. Booth, A. Bretaudeau, J. Brezovsky, R. Casadio, G. Cesareni, F. Coppens, M. Cornell, G. Cuccuru, K. Davidsen, G. D. Vedova, T. Dogan, O. Doppelt-Azeroual, L. Emery, E. Gasteiger, T. Gatter, T. Goldberg, M. Grosjean, B. Grüning, M. Helmer-Citterich, H. Ienasescu, V. Ioannidis, M. C. Jespersen, R. Jimenez, N. Juty, P. Juvan, M. Koch, C. Laibe, J.-W. Li, L. Licata, F. Mareuil, I. Mičetić, R. M. Friborg, S. Moretti, C. Morris, S. Möller, A. Nenadic, H. Peterson, G. Profiti, P. Rice, P. Romano, P. Roncaglia, R. Saidi, A. Schafferhans, V. Schwämmle, C. Smith, M. M. Sperotto, H. Stockinger, R. S. Vařeková, S. C. Tosatto, V. de la Torre, P. Uva, A. Via, G. Yachdav, F. Zambelli, G. Vriend, B. Rost, H. Parkinson, P. Løngreen, S. Brunak, Tools and data services registry: a community effort to document bioinformatics resources, Nucleic Acids Research 44 (2015) D38–D47. URL: https://doi.org/10.1093/nar/gkv1116. doi:10.1093/nar/gkv1116.

[37] W. Hugo, Y. Le Franc, G. Coen, J. Parland-von Essen, L. Bonino, D2.5 FAIR Semantics Recommendations Second Iteration, Project Deliverable D2.5, FAIRsFAIR, 2020. doi:10.5281/zenodo.4314320.

[38] British Oceanographic Data Centre, 2022, The NERC Vocabulary Server, URL: https://vocab.nerc.ac.uk/.

[39] R. F. d. Silva, L. Pottier, T. Coleman, E. Deelman, H. Casanova, Workflowhub: Community framework for enabling scientific workflow research and development, in: 2020 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS), 2020, pp. 49–56. doi:10.1109/WORKS51914.2020.00012.

[40] github, GitHub, 2022. URL: github.com.

[41] OpenAIRE, 2022, Argos, URL: argos.openaire.eu.

[42] M. Aristarán, M. Tigas, J. B. Merrill, 2022, Tabula, URL: ManuelAristarán, MikeTigasandJeremyB.Merrill.

[43] J. Graybeal, N. Vasiljevic, 2022, excel2rdf-template, URL: https://github.com/fair-data-collective/excel2rdf-template.

[44] University of Tübingen, 2022, Cmdi to dublin core transformer, URL: https://weblicht.sfs.uni-tuebingen.de/converter/Cmdi2DC/.

[45] Creative Commons, 2022, License chooser, URL: https://creativecommons.org/choose/.

[46] EUDAT, 2022, License selector, URL: https://github.com/ufal/public-license-selector.

[47] A. Devaraju, M. Mokrane, L. Cepinskas, R. Huber, P. Herterich, J. de Vries, V. Akerman, H. L'Hours, J. Davidson, M. Diepenbroek, From conceptualization to implementation: FAIR assessment of research data objects, Data Science Journal 20 (2021). doi:10.5334/dsj-2021-004.

[48] FAIRSharing, 2022, FAIRAssist, URL: fairassist.org.

[49] S.-A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A. L. Lister, M. Thurston, the FAIRsharing Community, Fairsharing as a community approach to standards, repositories and policies, Nature Biotechnology 37 (2019) 358–367. doi:10.1038/s41587-019-0080-8.

[50] N. A. Smale, K. Unsworth, G. Denyer, E. Magatova, D. Barr, A review of the history, advocacy and efficacy of data management plans, International Journal of Digital Curation 15 (2020). doi:10.2218/ijdc.v15i1.525.

[51] S. Jones, R. Pergl, R. Hooft, T. Miksa, R. Samors, J. Ungvari, R. I. Davis, T. Lee, Data management planning: How requirements and solutions are beginning to converge, Data Intelligence 2 (2020). doi:10.1162/dint_a_00043.

[52] H. P. Sustkova, K. M. Hettne, P. Wittenburg, A. Jacobsen, T. Kuhn, R. Pergl, J. Slifka, P. McQuilton, B. Magagna, S.-A. Sansone, M. Stocker, M. Imming, L. Lannom, M. Musen, E. Schultes, FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources, Data Intelligence 2 (2020) 158–170. doi:10.1162/dint_a_00038.

[53] Ministry of Education and Culture, Finland, 2022, Fairdata services, URL: fairdata.fi.

[54] E. Schultes, B. Magagna, K. M. Hettne, R. Pergl, M. Suchánek, T. Kuhn, Reusable FAIR implementation profiles as accelerators of FAIR convergence, in: G. Grossmann, S. Ram (Eds.), Advances in Conceptual Modeling, Springer International Publishing, Cham, 2020, pp. 138–147. doi:10.1007/978-3-030-65847-2_13.