FAIR OSF2022NL main text

Words: 4000-5000
Audience: data stewards / data experts / members of research community interested in FAIR/standardisation but not sure where to start
References: please add references to the Zotero library

Title: Ten simple rules for starting FAIR discussions in your community

Contributors in alphabetical order: Frédérique **Belliard**, Angelica **Maineri**, Esther **Plomp**, Andrés Felipe **Ramos Padilla,** Junzi **Sun,** Maryam **Zare Jeddi**

CRediT (Holcombe et al. 2020): Link to sheet

Keywords: FAIR principles, consensus standard, Community building, Curation workflow, domain-relevant community standards

Abstract

This work presents ten rules that provide guidance and recommendations on how to start up discussions around the implementation of the FAIR (Findable, Accessible, Interoperable, Reusable) principles and creation of standardised ways of working. These recommendations will be particularly relevant if you are unsure where to start, who to involve, what the benefits and barriers of standardisation are, and little work has been done in your discipline to standardise metadata or documentation methods. When applied, these rules will support a more effective way of engaging the community with discussions on standardisation and practical implementation of the FAIR principles.

Introduction

The FAIR data principles promote good data stewardship by leveraging the Findability, Accessibility, Interoperability, and Reusability (FAIR) of research data and software (Barket et al. 2022; Wilkinson et al. 2016, Lamprecht et al. 2020, Jacobsen et al. 2020). These principles aim to facilitate the discovery, access, integration, and reuse of research data and software by both humans and machines, with the ultimate goal of enhancing the transparency, reproducibility, interoperability, and impact of research. The FAIR principles emphasise the need for standardisation in the way research objects are described, stored, and shared: while FAIR is not *a* standard, the principles suggest that multiple FAIR standards should be developed (Mons et al. 2017 10.3233/ISU-170824 p. 51). By promoting consistent and well-defined data structures, controlled vocabularies, and metadata, the FAIR principles can help make research objects more easily comparable and reusable across different disciplinary and spatial contexts.

Despite the benefits of the FAIR principles and their widespread endorsement on behalf of research institutes, publishers, and funders, these principles have not been evenly adopted in all disciplines (Genova et al. 2017). There is still a lack of data and code sharing (with estimates between 1-20% (Federer et al. 2018; Hardwicke et al. 2020; Harris et al. 2018; Serghiou et al. 2021, French Open Science Monitor 2022) - although there are higher sharing rates in, for example, genomic research Anagnostou et al. 2015)). Furthermore, not every discipline has access to metadata standards or discipline specific repositories. One of the main challenges to the wider implementation of the FAIR principles is linked to the social dynamics underlying the standardisation process. Standardisation is a complex process that involves the creation of agreed-upon rules across time and space (Timmermans and Epstein 2010 10.1146/annurev.soc.012809.102629, p. 71). This process is difficult to facilitate without sufficient leadership, resources and time. Standardisation processes may also create frictions linked to a reduction of diversity, and to authority and governance issues (who decides on which standard to adopt?). These social dynamics are key to the successful implementation of the FAIR principles.

In this context, we shift our focus away from the specific research objects involved in standardisation processes and instead focus on the community aspect of standardisation. Specifically, we consider the strategies and approaches that can be employed to engage research communities in fruitful discussions about standardisation in the context of implementing the FAIR principles. In our view, the successful implementation of the FAIR principles relies on the buy-in and participation of the research community that will have to actually implement the principles.

To assist in the facilitation of standardisation discussions within individual research communities, we have developed the ten rules as a reference point (see **Figure 1** for an overview). Input on these rules has been initially provided by experts (including researchers, data supporters, students, service providers) at the Netherlands Open Science Festival on September 1st 2022 (see the Supplementary materials for more details) and a call for contributions via FAIR connect (Belliard 2023). It is important to note that not all research communities will be at the same stage of adoption of the FAIR principles, and some of these steps may be deemed unnecessary or irrelevant depending on the specific needs and circumstances of a given discipline. Our perspective will be biased towards the Dutch context as the Open Science Festival was hosted primarily for researchers in the Netherlands, and all authors are based at Dutch Institutes. Nevertheless, we hope that these rules serve as a useful resource for researchers, data supporters, and infrastructure providers, looking to effectively promote the adoption of the FAIR principles within their own communities.

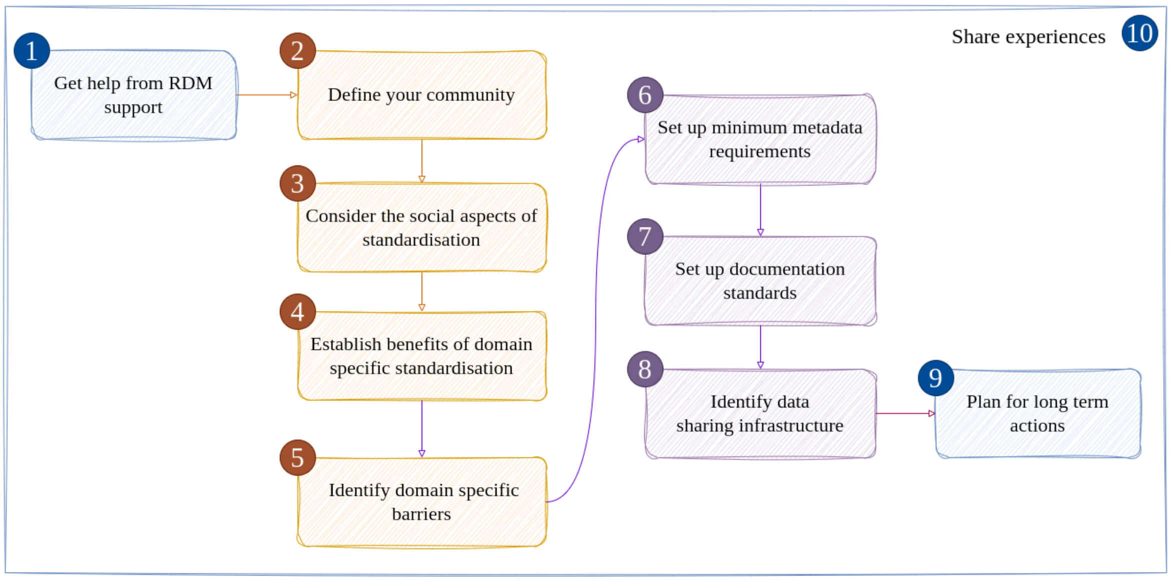

Figure 1. Overview of the ten rules to starting FAIR discussions. It is important to get support (Rule 1) and define the community that needs to be involved (Rule 2). In these discussions the social aspects of standardisation should be considered (Rule 3). It is therefore important to establish the benefits of domain specific standardisation (Rule 4) and to address the existing domain specific barriers (Rule 5). Keeping this in mind, it will become possible to set up minimum metadata requirements (Rule 6), documentation standards (Rule 7) and to identify the infrastructure that the community can make use of or should establish (Rule 8). In these efforts the long term sustainability of the efforts (Rule 9). For each of these steps it is important to share experiences (Rule 10).

Rules

Rule 1. Get help from RDM support at your institute or scientific association

Attempting data standardisation is a complex process that should not be done alone. Support is most likely to be found from the research data management (RDM) support team at your institute (usually based at the library), or the scientific association of your discipline (see **Table 1** for some examples). Failing to engage with these stakeholders may result in a lack of awareness and recognition of the need for standardisation in your discipline within your institute or association. We recommend prioritising this type of support, as it could prove to be beneficial in the long run, even if funding or resources may not be immediately available.

The RDM team will be able to point to existing resources, tools and information that can save time. They can also provide support in raising awareness, as they should have experience with promoting the adoption of FAIR principles within the research community. This team will likely have experience with providing workshops, training programmes, setting up policies and recommendations, and hosting events. They may have already set up materials or resources that can be tailored to specific needs.

The scientific association may have already set up data standardisation processes, will provide connections to your community (see **Rule 2**), and can provide further support.

Table 1. Examples of organisations that can help finding RDM experts grouped by spatial focus (primarily in Europe) and domain specificity. The examples are meant to give an indication, not an exhaustive overview.

Focus	**Domain-specificity**	**Examples**
Institutional level		(Digital) competence centres - university libraries - university research support team - institutional/faculty/department data steward Open Science Communities (see International Network of OSC)
National level	Domain-specific	Research Infrastructures, scientific associations: - ELIXIR's national nodes (life sciences) - Italy's Foster Open Science in Social Science Research (FOSSR) - the Dutch Open Data Infrastructure for the Social Science and Economic Innovations (ODISSEI)
	Domain-agnostic	Data archives, scientific associations: - National data archives (see for examples the list of contributors to the CESSDA data management guide or the Australian Data Archive - ADA) - Research Data Alliance (RDA) National groups - Research Data Access & Preservation (RDAP) RDM national initiatives: - the Netherland's LCRDM (see LCRDM's pool of experts) - the Netherland's Data Stewardship Interest Group (DSIG) - Italy's Italian Computing and Data Infrastructure (ICDI) Competence Center (link) - UK's Data curation center (DCC) - Denmark's National Forum for Research Data Management
International level	Domain-specific	Research Infrastructures: - Consortium of European Social Science Data Archives (CESSDA, https://www.cessda.eu/) - ELIXIR, a European infrastructure for the life sciences
	Domain-agnostic	Scientific associations and interest groups: - RDA's working and interest groups - RDA/EOSC Future ambassadors - GOFAIR Implementation Networks

Ideally, there is funding involved in standardisation efforts. Both RDM support and scientific associations may also have access to funding, or they may be able to connect you to other funding opportunities. Funding agencies in the Netherlands allow for funding data management activities in their projects (NWO - https://www.nwo.nl/en/research-data-management and ZonMW - https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/data-management-in-your-project/). Open Science Funds may be available through Open Science Communities (https://osc-delft.github.io/funding#mainstreaming-open-science-fund) or at a (inter)national level (https://www.nwo.nl/en/researchprogrammes/open-science/open-science-fund; https://fair-impact.eu/fair-impact-open-calls-support).

Once you have found support, the next step is to see who else needs to be involved (**Rule 2**).

Rule 2. Define the community you want to approach

Discussions surrounding the standardisation of FAIR practices should occur within a research community, defined as a group of stakeholders (such as individual researchers, data supporters, infrastructure managers, funders) that have a shared interest in streamlining their efforts to implement the FAIR principles. As explained by Timmermans and Epstein (2010 10.1146/annurev.soc.012809.102629), standardisation is inherently a social process that requires the commitment and endorsement of multiple actors to be effective. The community aspects of FAIR implementation are embedded in the original FAIR principles (Wilkinson et al 2016 10.1038/sdata.2016.18), and made explicit in principle R1.3 ("(meta)data meet domain-relevant community standards"). For instance, in the framework of FAIR Implementation Profiles (FIPs, a methodology that has been introduced to document FAIR implementation choices), the community aspect is captured by the concept of FAIR Implementation Community (Schultes et al. doi.org/10.1007/978-3-030-65847-2_13; Maineri and Wang 2022 https://doi.org/10.5281/zenodo.7428411). How to adequately define and engage a community remains, however, an open challenge.

Rule 2 recognizes that identifying the appropriate research community is a crucial step in facilitating discussions on standardisation. Research communities can be based on various factors, such as the type of data being generated or used, a shared institutional affiliation, or a specific research project. A community can constitute a formal entity, or it can be an informal group; and it can exist for a determined time span, or be long-lasting (Schultes et al. doi.org/10.1007/978-3-030-65847-2_13). It is important that a community self-identifies as such, as this can increase the level of commitment and engagement among members.

Research communities typically involve individuals in a variety of different roles, such as researchers, RDM support staff, lab technicians and students. Once the community has been identified, the levels of understanding of FAIR implementation and FAIR standards should be gauged, using resources such as the FAIR-Aware tool developed by DANS (Akerman et al 2021 https://doi.org/10.5281/zenodo.5084861) or the How to FAIR quiz from the Danish National Forum for Research Data Management (Deutz et al 2020 https://doi.org/10.5281/zenodo.3712065). Disparities in understanding among different stakeholders may present challenges to the standardisation process - though a diversity in perspectives can be beneficial, as elaborated in **Rule 3.**

Depending on the features of the community, there may be different ways to get in touch with the community members: in the case of an informal community, for instance, it may be necessary to proceed via "snowballing", with one identified member suggesting other ones and so on; in the case of formalised communities, instead, there may be people with specific roles (such as community managers) who already have open communication channels with the community. In both cases, reaching out to RDM experts, research infrastructures or scientific associations may be beneficial (see **Rule 1**) as they may be aware of existing or similar initiatives, or be able to suggest people to contact.

Rule 3. Consider the social aspects of standardisation

As discussed in **Rule 2**, standardisation is a social process that requires the participation and commitment of multiple actors within a research community (Timmermans and Epstein, 2010 10.1146/annurev.soc.012809.102629). Standardisation efforts can stop or become obsolete when a community loses interest (Sansone and Rocca-Serra 2016 https://doi.org/10.6084/m9.figshare.4055496.v1). Therefore, **Rule 3** highlights the importance of considering the social aspects that may influence the consensus-building process when facilitating discussions on the standardisation of FAIR practices and offers practical examples.

First of all, the creation and adoption of standards may involve trust and authority issues. For example, the introduction of new standards may generate opposition and resistance if community members do not perceive them as legitimate or do not trust the authority of those proposing the standards. To mitigate the risk of frictions, involving the community is key. In particular, highlighting best practices via real use cases from the community can be effective in showing the value of standardised FAIR practices and promoting the use of existing standards.

When there are no existing best practices yet within your own community, examples and recommendations can be adopted from other communities who are further along in adopting the FAIR principles (see **Rule 1** - RDM experts can help reaching out to more mature communities; but also see **Rule 10**: sharing experiences is pivotal!). Building on existing practices may save time and resources that can be used more efficiently in the standardisation process.

Additionally, standardisation can sometimes result in a reduction of diversity within a community, as it involves the creation of agreed-upon rules that may limit the range of options available. Therefore, it is important to carefully consider the potential impacts of standardisation on the diversity of research practices within a community. The community-based approach to the FAIR principles, alongside the FIPs which support the documentation of FAIR implementation choices (Maineri and Wang, 2022 10.5281/zenodo.7428412), make it possible to create multiple standards, as long as cross-standard interoperability is kept into account. For instance, the CEDAR Metadata Tools allow the creation of community-specific metadata templates while reusing existing ontologies and value sets, therefore enabling diverse solutions within a shared framework (see **Rule 6** for more recommendations on creating metadata models). There are also solutions to make existing standards more easily findable and reusable.One such solution is FAIRsharing which serves as a repository of FAIR enabling standards and other FAIR resources. This platform was created to address the issue of excessive fragmentation in the developments of standards (Sansone et al. 2019 10.1038/s41587-019-0080-8).

Standardisation often requires researchers to invest time and resources into changing their existing workflows, which can be a challenging task. The disruption of existing practices has also been reported to be an important barrier to adoption of standards in the industry context (see Storz 2007 10.1093/ssjj/jym048) (see also **Rule 5**). Therefore, it is important to clearly communicate the benefits and incentives of adopting the FAIR principles to the research community. By clearly communicating the value of the FAIR principles and engaging in meaningful dialogue with the research community, it is possible to facilitate a more effective and efficient standardisation process (see also **Rule 4**).

Rule 4. Establish benefits of domain specific standardisation

It can be helpful to establish benefits of standardisation in order to convince others to get involved in domain specific standardisation. Some examples of benefits are listed below, and may not always be directly applicable to your community.

Personal benefits by being involved in this process include:

- Direct impact on the eventual results, ensuring that the standardisation processes are applicable to your research.
- Extension of professional network and positive effects on your professional reputation.
- Long term, standardisation can be more cost efficient by increasing data reuse (Pronk 2019), preventing duplication of research/trials, improving quality of the data (reduction of data errors and increase of reproducibility) and facilitating integration of datasets.

Domain specific standardisation also reflects positively on research funders and institutions (see **Rule 1** on where to find help in your institute). Benefits for institutions include:

- Streamlining data management processes and being cost-efficient.
- The reputation and trust in research findings from the related research groups (Sequiera 2021).
- Rewards and recognition of data as a valuable research output in academia.
- Facilitating collaboration and innovation (Sequiera 2021).
- Increasing the value of existing data by facilitating reuse, and thereby increasing the return on the initial investment on data collection.

After the benefits have been established, the next step is to identify the barriers (**Rule 5**).

Rule 5. Identify domain specific barriers

It is important to recognize that the barriers and challenges to implementing the FAIR principles and standardising data management practices may vary widely across different research domains. This may involve performing a gap analysis to identify areas where additional support or resources are needed, or identifying case studies of successful FAIR implementation in similar disciplines that can serve as best practice for the community (see also **Rule 3**). Below are some of the barriers that we have encountered, either in the literature or in our own professional experience:

- **Data requirements**: the types of data more commonly used by a community as well as their size and heterogenous nature may pose challenges in terms of storage, management, and accessibility (Dallmeier-Tiessen et al. 2014; Gomes et al. 2022; Perrier et al. 2020).
- **Ethical and legal barriers**: The handling and sharing of sensitive political or personal data may be subject to strict regulations (such as the General Data

Protection Regulation (GDPR)) and require additional considerations to ensure compliance (Ali-Khan et al. 2017; Dallmeier-Tiesen et al. 2014; Devriendt et al. 2021; Fecher et al. 2015; Gomes et al. 2022; Harris et al. 2018; Perrier et al. 2020; Van Panhuis et al. 2014). Data sharing can also result in economic damage when disease data is shared and impacts tourism and trade (Van Panhuis et al. 2014).

- **Different research environments**: When there are critical resource shortages (such as the absence of research networks and lack of infrastructural support) there may be more immediate concerns that should be addressed (Bezuidenhout & Chakauya 2018)
- **Intellectual property and licensing**: Intellectual property (IP) issues (such as data transfer and processing agreements) may arise when data are shared or reused (Gomes et al. 2022; Harris et al. 2018; Perrier et al. 2020), particularly when multiple stakeholders are involved. Not everyone may have access to proprietary software used in data analyses.
- **Lack of incentives**: Some researchers may not see the value in making their data FAIR or may not perceive a need to share their data with others (see **Rule 3**) Ali-Khan et al. 2017; Bezuidenhout 2018; Chawinga and Zinn 2019; Dallmeier-Tiesen et al. 2014; Devriendt et al. 2021; Gomes et al. 2022; Fecher et al. 2015; Poline et al. 2022; Van Panhuis et al. 2014).
- **Cultural barriers**, for example, considering data sharing to hamper future publications if there is no reciprocity in the form of appropriate credit for data sharing (Ali-Khan et al. 2017; Bezuidenhout 2018; Perrier et al. 2020; Van Panhuis et al. 2014), or a lack of trust in data being correctly interpreted and used (Chawinga and Zinn 2019; Dallmeier-Tiesen et al. 2014; Devriendt et al. 2021; Gomes et al. 2022; Fecher et al. 2015; Van Panhuis et al. 2014).
- **Lack of Institutional data policy, support and training** (Borghi and VanGulick 2021).
- **Lack of infrastructure** (see **Rule 8**) (Bezuidenhout 2018; Chawinga and Zinn 2019; Dallmeier-Tiesen et al. 2014; Perrier et al. 2020; Van Panhuis et al. 2014).
- **Lack of compliance monitoring** by institutes, funders or journals (Anger et al. 2022; Chawinga and Zinn 2019; Dallmeier-Tiesen et al. 2014; Harris et al. 2018).
- **Limited awareness** about best practices, FAIR principles and standards (Fecher et al. 2015).
- **Emphasis on novel research** may result in data generation rather than reuse, integration and maintenance (Poline et al. 2022).
- **Possible criticism** and fear that results will be invalidated (Dallmeier-Tiessen et al. 2014; Van Panhuis et al. 2014).
- **Limited time and/or lack of resources** (Ali-Khan et al. 2017; Borghi and VanGulick 2021; Chawinga and Zinn 2019; Dallmeier-Tiesen et al. 2014; Devriendt et al. 2021; Fecher et al. 2015; Harris et al. 2018; Perrier et al. 2020; Van Panhuis et al. 2014).

To identify and address these barriers, you should discuss them with your community of stakeholders (**Rule 2**).

Pending on the identified barriers there will be different solutions to address them. Some of these barriers (limited awareness, lack of expertise and best practices) can be addressed by data standardisation and defining more explicitly what information is most valuable in the data management workflows. In **Rule 6** this is addressed by going deeper into how a metadata standard can be established.

Rule 6. Set up minimum metadata requirements

Metadata is information about the data that provides context and allows for proper interpretation and reuse. A metadata standard is a structured form of documenting and describing this information. Several metadata standards are already in use, such as Dublin Core and DataCite (FAIRsharing Team 2018; see the Digital Curation Center for an overview of metadata standards, or use FAIRsharing to browse metadata standards). Dublin Core consists of fifteen general elements that makes this standard easy to use across disciplines. Nevertheless, most disciplines may require more detailed metadata than those provided by Dublin Core. It is therefore helpful to look for discipline specific metadata standards or guidelines (see FAIRsharing.org (Sansone et al. 2019) or the Metadata Standards Catalog). When there are no metadata standards or minimum metadata requirements available for your discipline, a more advanced step is to start creating these minimal metadata requirements. This can be a complex task, particularly when the research discipline overlaps with different fields that use distinctive terminology to describe data (Scheffler et al. 2022).

To start with setting up minimum metadata requirements it is important to first establish who needs to be involved as community engagement will be crucial (Scheffler et al. 2022) (**Rule 2**). Other communities have already been successful in developing minimum metadata requirements, such as the Earth Sciences (Crystal-Ornelas 2022; Slaton et al. 2021), Bionano Sciences (Faria et al. 2018), Biomedical field (Sarkans et al. 2021) and -omics Sciences (Fiehn et al. 2007; Kolker et al. 2014; Perez-Riverol et al. 2020; Sumner et al. 2007). The lessons learned from these communities can be taken into account, although their approaches may not always fit with your research community that may have different requirements and challenges (**Rule 5**).

Minimum metadata requirements can be about data/sample pre-processing, experimental analysis, quality control, preregistration - any aspect related to the research process. The metadata requirements should provide guidelines for essential information requirements, while at the same time be flexible to meet each researcher's objectives (Mészáros et al. 2022; Sarkans et al. 2021). There is a need to 'strike the right balance between minimising the barriers to data submission and maximising opportunities for data reuse' (Sarkans et al. 2021).

After you identified the relevant stakeholders (**Rule 2**), you can follow the recommendations below to start setting up minimum metadata requirements, or a Minimum Information Standard, in a research community:

- Review existing practices such as metadata standards, guidelines and use cases (Crystal-Ornelas 2022; Fiehn et al. 2007; Kolker et al. 2014; Mészáros et al. 2022; Slaton et al. 2021).
- If there are no existing efforts, you can start with a call for guidelines (see Perez-Riverol et al. 2020), or set up a working group/project team (Sumner et al. 2007; Slaton et al. 2021) or a network (Hollman et al. 2021). You can get a team together by organising a workshop or conference session (Fiehn et al. 2007;

Mattingly et al 2016; Sarkans et al. 2021; Vardigan et al. 2008). Ideally, funding is available for standardisation efforts (for example, NIH funding / Sansone and Rocca Serra 2016), or should be applied for (Hollman et al. 2021; Mattingly et al. 2016) (see also **Rule 1**).

- Adhering to the minimum metadata requirements should be as effortless as possible to enable widespread adoption (Mattingly et al 2016; Mészáros et al. 2022) (see also **Rules 3** and **4**).
- Minimum metadata requirements are the first step towards standardisation. Additional developments will be needed for standardisation (Fiehn et al. 2007; Mészáros et al. 2022), involving the research community at each step.
- To establish community consensus, the research community should be asked for input and feedback, through community discussions, workshops and surveys (Crystal-Ornelas 2022; Fiehn et al. 2007; Kolker et al. 2014; Mattingly et al. 2016; Perez-Riverol et al. 2020; Sarkans et al. 2021). Only through active community involvement will a functional solution be achieved (Sumner et al. 2007) (see also **Rule 2**).
- To ensure practical and effortless implementation of the standards by journal editors, reviewers, and data repositories, it is important to gather their feedback (Sumner et al. 2007; Slaton et al. 2021; Sarkans et al. 2021; Vardigan et al. 2008).
- Once progression has been made, it is important to communicate this to the research community, via public documentation, reports or publications (see also **Rule 10** on sharing experiences).
- To support community uptake it can be helpful to provide training, support or to have champions involved that can promote the standards (Crystal-Ornelas 2022). The benefits (**Rule 4**) of the uptake of the metadata standard should be clear.

A great way to get started is to review the work by ESS-DIVE in establishing a community-centric metadata reporting format (Crystal-Ornelas 2022). Crystal-Ornelas et al. share guidelines (Box 1) and details about their process. A next step could be to develop an ontology (see Courtot et al. 2016's ten simple rules on this topic).

Rule 7. Set up documentation standards

In addition to the minimum metadata requirements described earlier (**Rule 6**), documentation is the next step that reports research outputs to allow/enable its reuse. Communities should be aware of the two levels of documentation that can be considered when documenting research data: project-level and data-level (also referred to as study-level and object-level documentation, respectively). The project-level documentation provides context for the collection, methodology, structure and validation of data, while the data-level documentation consists of the variable names, descriptions, classifications, file formats and software details. In other words, at project-level documentation is about what is around data and at data-level documentation is about data itself (CESSDA Training Team 2020).

These two levels of documentation require different standardizations. Examples of project level documentation are Data Management Plans (DMPs), Software Management Plans (SMPs) and the use of Preregistration and Registered Reports. The first two provide project-level documentation by describing the context of data and software; in DMPs describing how data was collected and the methods used to validate it (The Turing Way 2023); in SMPs, describing how software works, the purpose, the outputs and its (continuous) development. Research communities may select standard templates for DMPs, taking into account the requirements of their organisation or funder (see for examples the list of public templates on DMPTool), while for SMPs such standards are still under development (Martinez-Ortiz et al. 2022). Preregistration on the other hand, involves the public disclosure of research plans before data collection, analysis, and reporting are completed, with the goal of increasing transparency in the knowledge creation process from its inception to the results (Evans TR. et al., 2023; Hardwicke TE and Wagenmakers EJ, 2023). In an effort to standardise the information requested for pre-registering a study and ideally simplify the process for researchers, templates have been developed (see the list of templates on the Open Science Framework website).

At the data-level, once the minimum metadata requirements are established (**Rule 6**), it will be easier to describe variables or file formats that the community will use and then expand to documentation guidelines. Documentation can be used to describe how to organise data, such as spreadsheets (Broman and Woo 2018), workflows (see the Data Curation Network Primers, 2023), and to provide information on standards for dates and times (such as ISO 8601 or RFC3339). In addition, it may be possible to use documentation from other communities. In particular, code programming communities use standard style guides that are also widely used within research communities (such as PEP 8 for Python, Tidyverse for R, see also guidance by the Code Refinery (2023)).

A recommended solution for documentation could be codebooks, which describe the variables with their units, summarising choices made during the research process, and outlining the experimental study design (Ellis and Leek 2018). Ideally, codebooks should be in a structured/standard format (for example, Data Documentation Initiative Codebook). Recently, tools have been developed that can automatically generate standardised metadata, reducing the (time) barriers to writing comprehensive codebooks (codebook R Package, Arslan 2019).

Ultimately, the choice of documentation standardisation should facilitate communication and collaboration between researchers and those who reuse their data. These instruments can be effective in reducing the workload in the medium to long term by enabling the incorporation of FAIR standards early in a research project.

Rule 8. Identify infrastructure to share data

In order to get a clear idea of the infrastructure that can be used to share data, first it is important to follow the requirements and guidelines of your institution, funders and/or collaborators. Generally, data repositories are considered to be the ideal infrastructure to share data in a reliable manner (ref). Generic or institutional repositories, such as Zenodo, OSF, or Figshare, are widely used for preserving and sharing research data (see **Table 2** for some examples).

Table 2. A list of common repositories outlined in more detail in Stall (2022).

Repository	**Dataset limit**	**Host**	**Comments**	**Certification**
Zenodo	50 GB	CERN (CH)	General-purpose repository	No
Figshare	20 GB 5 TB (Figshare+)	Digital Science (UK)	General-purpose repository	ISO27001
Harvard Dataverse	1000 GB	Harvard University (US)	General-purpose repository	No
Dryad	300 GB	Dryad Digital Repository, Inc. (US)	Popular for research data underlying scientific and medical publications	CoreTrustSeal
4TU.Research Data	100 GB	4TU.Centre for Research Data (NL)	Popular for technical and scientific research data	CoreTrustSeal
Open Science Framework	50 GB (public) 5 GB (private)	Center for Open Science (US)	Information and data sharing platform	No
Mendeley Data	10 GB	Elsevier (NL)	General-purpose repository	No

Discipline specific infrastructure is especially beneficial if standard data formats are used and enforced, ideally via user friendly interfaces and with training provided where needed (Mayer et al. 2021). There are many domain-specific repositories (see NIH list of domain-specific repositories). You will need to verify that these repositories follow the minimum requirements to be considered useful and FAIR for research practices. Think that these repositories need to:

1) Have a clear policy on how data will be managed, as well as a privacy policy and a terms of use
2) Provide sufficient data storage size for your dataset.
3) The geographic location where the data is saved (for restricted-access dataset that contains personal data).
4) Assign a persistent identifier (such as Digital Object Identifier, DOI) to be able to cite the data
5) Allow you to include a licence to your data (such as a Creative Commons licence).
6) Make sure data is available/accessible and discoverable. Repositories can enhance their discoverability by being included in databases such as re3data (https://www.re3data.org, Pampel et al. 2013), FAIRsharing (https://fairsharing.org, Sansone et al. 2019), and the EOSC portal (https://eosc-portal.eu).
7) Allow revisions to be made to the dataset in the future.

In some cases where an institutional or generic repository does not fulfil the desires or challenges of your community (see **Rule 5**), setting up a specific (discipline or project) repository may be a good option when there are plans for the long-term sustainability of the infrastructure (see **Rule 9**). The main advantage of arranging your own repository infrastructure is that you have greater control over how data is documented and presented to the public and/or researchers. Specific data repository infrastructure may also improve the data quality of the datasets (Kindling and Strecker 2022). However, creating and using this infrastructure leads to additional costs (especially when dealing with large quantities of data). In addition, clear documentation (see **Rule 7**) and training materials (see **Rule 9**) are required to engage researchers using the repository.

Rule 9. Plan for the long term actions

As mentioned in **Rule 8**, discipline specific infrastructure requires resources and should be sustainable for the long-term. When these infrastructures are set up by a small group, maintenance and sustainability is challenging as many researchers move across institutes and countries. While sustainability can be achieved by charging for repository services, it is also important to consider that not all researchers have access to these resources. Researchers are generally working on projects that eventually run out of funding, especially at the stage of data sharing. It is therefore important to consider who pays for long term data sharing and maintenance. Maintenance plans and governance of infrastructure and standards should be transparently communicated (see also **Rule 3**).

To plan for the long-term and to establish robust policies for the repository, repositories could aim for certification (via CoreTrustSeal, ISO 16363 or Nestor), although this is a resource intensive process. Resources are also needed for the maintenance of any created metadata standards (**Rule 6**). Standardisation is a continuous process and will require evaluation on their practical applicability (for example, metadata standards may become obsolete or deprecated when they are no longer applicable - Sansone and Rocca-Serra 2016). Standardisation is also a continuous learning process. Researchers may not be familiar with the standardisation efforts and will need a place to start, support or training resources. It can also be important to monitor whether standardisations are followed appropriately - some form of manual curation may always be needed to avoid errors or incomplete entries. All of these processes take up resources in the long term.

To facilitate long-term sustainability it is better to use open formats and infrastructure built using open source software. This prevents lock-in to certain services and allows community members to continuously contribute. It is also important to consider how the infrastructure will scale when future use is increasing and user input may become more heterogeneous (Klump et al. 2023).

To foster a culture of data standardisation and sharing it is needed to recognise the efforts of researchers who adopt minimum metadata requirements (**Rule 4**). Ideally this happens at the institutional level. Research communities can also recognise practices during annual meetings and conferences, or awarding prizes (for example, the Open Science Community Amsterdam Awards that took place on the 26th of January, OSCA 2023).

Rule 10. Share experiences

Sharing experiences about the standardisation process facilitates learning from existing efforts and identifying best practices that can facilitate the standardisation journey. By reaching out to local RDM support or other community stakeholders (see **Rule 1**), you have hopefully benefitted of the experiences of others as well. It is therefore important to share experiences gained from each of the rules listed here, and as illustrated in **Figure 1**. Experiences and insights can be shared via case studies, best practices, and lessons learned from standardisation efforts. Venues to share these experiences may include journals (such as PLOS), preprint servers, publishing forums (such as FAIR connect, where an earlier version of this article was shared - Belliard et al. 2023), data repositories (such as Zenodo), blogs (for example, Kalvera 2023), social media, or conferences and meetings.

Conclusion

Standardisation efforts are complex and time consuming. The recommendations that we share emphasise the need to find support (**Rule 1**) and identify the relevant stakeholders that need to be involved (**Rule 2**). As standardisation efforts are primarily a social issue (**Rule 3**), it is important to identify the benefits (**Rule 4**) and to address the existing barriers (**Rule 5**). Keeping this in mind, it will become possible to set up minimum metadata requirements (**Rule 6**), documentation standards (**Rule 7**) and identify the infrastructure that the community can make use of or should establish (**Rule 8**). It is important for infrastructure, but also standardised procedures, to consider the long term sustainability of the efforts (**Rule 9**). Crucial to each of these steps is the sharing of the lessons learned and the materials created, so that others do not have to start from scratch (**Rule 10**). By following these recommendations, you should be able to more successfully engage your community in standardisation discussions, resulting in successful implementations of the FAIR principles.

Acknowledgements

We are grateful to the over 40 participants at the Session “*Starting FAIR discussions increasing standardisation in your research community*” at the Open Science Festival in Amsterdam on September 1st, 2022 for their initial input on the checklist (see the Supplemental Materials). Thanks to FAIRconnect for helpful comments by Erik Schultes and Barbara Magagna.

CRediT

Author contributions set up via Tenzing (Holcombe 2020):

Frédérique Belliard: Conceptualization, Writing - original draft, and Writing - review & editing.
Angelica M. Maineri: Conceptualization, Writing - original draft, and Writing - review & editing.
Esther Plomp: Conceptualization, Project administration, Supervision, Writing - original draft, and Writing - review & editing.
Andrés F. Ramos Padilla: Writing - original draft and Writing - review & editing.
Junzi Sun: Conceptualization, Visualization, Writing - original draft, and Writing - review & editing.
Maryam Zare Jeddi: Writing - review & editing.

References

Please add references to the Zotero library.

Supplemental Materials

Process to getting to ‘*Ten simple rules for starting FAIR discussions in your community*’

Starting FAIR discussions in your community was originally conceived as a session for the Open Science Festival 2022 by Esther Plomp. After internal discussions at TU Delft between Esther and Frédérique, and a call for contributions on Twitter, the session was submitted by 13 May 2022 (see **Table S1** for the proposal text).

Table S1: Session proposal submitted to the Open Science Festival 2022.

Information fields of the submission form	**Answers by the contributors of the session proposal**
Names of organiser(s):	Esther Plomp, Delft University of Technology - Faculty of Applied Sciences Frédérique Belliard, Delft University of Technology - Library Junzi Sun, Delft University of Technology - Faculty of Aerospace Engineering Esther Maassen, Tilburg University - Faculty of Social and Behavioral Sciences Angelica Maineri, Erasmus University Rotterdam - Erasmus School of Social and Behavioural Sciences
Names of moderators / speakers if applicable;	Moderators: All organisers, pending on how many participants will join the session to form smaller groups.
please indicate if you want to organize a workshop or session:	Session (45 min)
Short description that summarizes the workshop / session, its rationale and relevance for the participants of the National Open Science Festival;	For data to align with the FAIR principles it has to follow domain-relevant standards (principle R1.3). Nevertheless, not all disciplines have metadata standards or standardised workflows, and the knowledge thereof varies considerably across different stakeholders. This session aims to exchange experiences on increasing standardisation in one’s research field. These experiences and recommendations will be translated to a short guide/checklist for anyone that wants to start standardisation discussions within their research community. Questions/steps to provide information on could be ‘how do we engage all the stakeholders?’ and ‘What would be the best way to structure/start these types of discussions?’ or ‘Is it possible to obtain funding for this?’ This aligns well with the festival’s proposed topics of: - Putting FAIR and open data into practice - Setting up and maintaining open communities - Working-sprints with peers on a specific open topic
Format and structure;	This session will be hosted as a discussion session in which participants will be split into smaller groups to brainstorm and gather input. These inputs will be gathered collectively. Interest will be gathered during the session for any follow up steps at the end. xx:00: Introduction to the topic and determining level of experience with the topic xx:05: Grouping of individuals with different experiences into smaller

	subgroups so that they can discuss xx:25: Short report out of the progress xx:30: Gathering initial checklists/steps and repeated points xx:40: Next steps xx:45: End of session
Target audience;	- Researchers that would like to start standardisation discussions in their own discipline. - Research professionals that would like to facilitate standardisation discussions. - Representatives of professional societies, data repositories, or funding agencies, and editors
Preferred outcomes for participants of the workshop / session;	Participants will build up their network and get to know individuals with a similar interest in data standardisation Start of a short guide/checklist which will be shared on a data repository once it is finalised (with contributors credited for their work)
Requirements for the workshop / session (screens, set up of tables, flipcharts, etc.)	Flipcharts Tables set up in small groups for discussions

Due to a role change, Esther Maassen could unfortunately no longer dedicate time to the session.

The session was accepted and advertised on the Open Science Festival 2022 website as following:

> **1 September 13:45 – 14:30**
>
> **Title: Starting FAIR discussions: increasing standardisation in your research community**
>
> **Abstract:** This session aims to exchange experiences on increasing standardisation in one's research field. These experiences and recommendations will be translated to a short guide/checklist for anyone that wants to start standardisation discussions within their research community. Questions/steps to provide information on could be 'how do we engage all the stakeholders?' and 'What would be the best way to structure/start these types of discussions?' or 'Is it possible to obtain funding for this?'

During the session preparations the schedule for the workshop was adjusted as following:

- xx:00: Introduction to the topic and determining level of experience with the topic
- xx:05: Subgroups (data collection, standardisation in publishing, community engagement, data sharing)
- xx:20: Switch from problem reporting to steps to take/solutions
- xx:35: Post stickies
- xx:40: Next steps (invite them to session on 8th of September)
- xx:45: End of session

The session was led by Frédérique Beillard, and facilitated by Junzi Sun and Angelica Maineri. After a brief introduction on the aims of the session (to co-create a checklist for anyone who wants to start FAIR standardisation discussions within their research community) the participants at the session split into five round tables. Four of these tables had predefined topics identified by the session organisers: data collection; standardisation in publishing; community engagement; and data sharing. The fifth table was suggested by the audience; Reusability. Recommendations developed at the tables were then shared via post its on a board. Participants were asked whether they were interested in further contributing to the checklist/10 rules via an online meeting taking place a week after the festival on the 8th of September.

Frederique C. Belliard @fredbelliard · Sep 1, 2022

Replying to @PhDToothFAIRy

Results of the full session on #FAIR standardisation **discussions.** Thanks @Junzi @AngelicaMaineri. @PhDToothFAIRy you were missed

Tweet by Frederique Belliard.

After the meeting on the 8th of September the group was extended with Andrés Felipe Ramos Padilla and Maryam Zare Jeddi. We discussed the session input (see the overview of the contributed stickies below). We started meeting regularly to work on the article via optional online co-working calls and discussion sessions until submitting the article in April 2023.

Summary of the stickies from the OSF session

Community engagement

- Define what community type
    - *In the community engagement table, some elements were mentioned that "make" a community: self-identification; sharing data types; institution*
- Community consensus on the path of data sharing
- Community momentum on creating domain specific metadata and publishing standards
    - *In the community engagement table, a point was raised that it can be effective to learn from communities that are further along in the FAIR standardisation process (Use cases?)*
- Policies for the institutes/projects about data sharing. For example, what to do when the owner/researcher goes away.

Data ownership could be transparent so it is not lost
standard in data usage policies
how to automate usage policies on infrastructure

- how to control policy violations

Storage platform - unified? Or everyone keeps their data?
What are the terms of use - licenses should be clear

Data sharing: standardisation, also documentation

Problems

- data sustainability: what happens at the end of the funding/project
- 'My data isn't good enough (yet) for sharing'
- Terms of use unclear
- metadata: community-specific requirements

Data sharing

- long-term planning: who is responsible (ownership), who will pay for keeping the data available?
- Where to share data/publish? Generic, discipline specific?
- Discipline specific documentation:
    - what information to provide to future (re)users
    - how to name your variables?
    - how to organize your files?
    - how can others (re)use data what can they do with it?
    - how were the data created

Follow the FAIR principles -> make the data as FAIR as possible

create a checklist for starting a conversation

- metadata catalog
- dmp
- standardisation vs harmonisation?
- ontologies
- it should not constrain science
- clear definitions of words. know your synonyms
- emphasize the benefits

- if there is no standard, document all your steps

Reward standardisation

standardised discipline-specific metadata requirements
discipline specific standardised vocabularies
sustainability

find the smallest possible standard

Think of different standards for different data types
Where/which platform
metadata

Where to share/publish your data -> can that be standardized in a community

Use open formats / Metadata standards / Preregistration / licenses / Open when possible

Minimum metadata requirement / Dublin Core

Documentation (readme file, metadata, code book)

File structure

Policy, standardisation vocabulary, procedure, metadata sheet

Certification of trusted repositories

Reusability

Advantages of data reuse:
- saves time
- more efficient data collection
- get more context/inspiration
- unexpected application / serendipity

rich metadata -> more reusability

Obstacles to data reuse
- existing data not (always) FAIR
- findability
- culture
- diversity/complexity = no interoperability

Ways forward / checklist
Reuse
- clear/complete metadata - domain-specific standardisation
- Create synthetic datasets if original data is sensitive
- codebooks (= a legenda of the dataset): standardised, machine-readable