Title: FAIRness Literacy: the Achilles' Heel of Applying FAIR Principles

Authors / Affiliations:

- Romain David, MISTEA, INRAE, Montpellier SupAgro, Université de Montpellier, Montpellier, France; david.romain@gmail.com; ORCID: 0000-0003-4073-7456;
- Laurence Mabile, INSERM-Université Paul Sabatier Toulouse III, Toulouse, France; laurence.mabile@univ-tlse3.fr; ORCID: 0000-0002-7724-1721;
- Alison Specht, SEES-TERN, the University of Queensland, Brisbane, Australia a.specht@uq.edu.au; ORCID: 0000-0002-2623-0854;
- Sarah Stryeck, Graz University of Technology, Institute for Interactive Systems and Data Science, Graz, Austria sarah.stryeck@tugraz.at; ORCID: 0000-0001-7524-6038;
- Mogens Thomsen, INSERM-Université Paul Sabatier Toulouse III, Toulouse, France; mogens.thomsen@univ-toulouse.fr; ORCID: 0000-0002-4546-0129;
- Mohamed Yahia, INIST-CNRS, Nancy, France; mohamed.yahia@inist.fr; ORCID: 0000-0001-7829-1918;
- Laurent Dollé, Managing Director Biothèque Wallonie Bruxelles, President BBMRI.be, Erasme Hospital, Brussels, Belgium; laurent.dolle@erasme.ulb.ac.be; ORCID:0000-0003-4566-6407;
- Clement Jonquet, LIRMM, University of Montpellier, CNRS, Montpellier, France; jonquet@lirmm.fr; ORCID: 0000-0002-2404-1582;
- Daniele Bailo, EPOS-ERIC / Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy; daniele.bailo@ingv.it; ORCID: 0000-0003-0695-4406;
- Elena Bravo, Istituto Superiore Sanità, Rome, Italy; elena.bravo@iss.it; ORCID: 0000-0002-3473-6767;
- Sophie Gachet, Aix Marseille Univ, CNRS, Avignon Université, IRD, IMBE, Marseille, France; sophie.gachet@imbe.fr; ORCID: 0000-0002-3599-5189;
- Hannah Gunderman, Carnegie Mellon University Pittsburgh, United States; hgunderm@andrew.cmu.edu; ORCID: 0000-0002-7710-7055;
- Jean-Eudes Hollebecq, MISTEA, INRAE, Montpellier SupAgro, Université de Montpellier, Montpellier, France; jean-eudes.hollebecq@inrae.fr; ORCID: 0000-0002-5089-9179;
- Vassilios Ioannidis, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; vassilios.ioannidis@sib.swiss; ORCID: 0000-0002-4209-2578;
- Daniel Jacob, INRAE, UMR BFP, Université de Bordeaux, Villenave d'Ornon, France; daniel.jacob@inrae.fr; ORCID: 0000-0002-6687-7169;
- Yvan Le Bras, DGD-REVE, UMS PatriNat, MNHN, Paris, France; yvan.le-bras@mnhn.fr; ORCID: 0000-0002-8504-068X;
- Emilie Lerigoleur, CNRS, UMR GEODE, Université Toulouse 2, Toulouse, France; emilie.lerigoleur@univ-tlse2.fr; ORCID: 0000-0002-0864-659X;
- Anne Cambon-Thomsen, INSERM-Université Paul Sabatier Toulouse III, Toulouse, France; anne.cambon-thomsen@univ-tlse3.fr; ORCID: 0000-0001-8793-3644;
- and the Research Data Alliance - *SHAring Reward & Credit (SHARC)* Interest Group https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig

Corresponding author's email address: david.romain@gmail.com

Authors' contributions

Writing Paper: RD, LM, SS, LD, DJ, YLB, CJ
Reviewing paper: ACT, JEH, MT, MY, SG, AS, HG, EB, EL,VI
Discussions about criteria implementation processes: SHARC IG members, especially during RDA 13th and 14th plenary meeting.
Key criteria conception: RD, LM, ACT, MT, MY
Survey conception: RD, LM
Survey review: SG, AS, ACT, MT, MY
Survey completion: JEH, DB, HG, VI,...
Survey analyses: RD

Abstract

The SHAring Rewards and Credit (SHARC) interest group was established in 2017 as part of the Research Data Alliance (RDA). The objective is to improve research crediting and rewarding mechanisms for scientists who strive to organise their data (and material resources) for community sharing. This implies that data are findable and accessible on the Web, and comply with shared standards making them interoperable and reusable. This was formalised in the FAIR principles in 2016 (Findable, Accessible, Interoperable and Reusable). Sharing requires considerable time, energy, expertise and motivation. One solution to encourage scientists to share their data would be to reward sharing as a first-class research output. To that aim, assessing the compliance of data with FAIR principles and increasing the understanding of FAIRness criteria – i.e. promoting FAIRness literacy, are critical steps. Appropriate human-readable criteria must first be identified in the FAIRness assessment processes. This paper presents the design and critical evaluation of a generic tool developed by the SHARC IG for assessing FAIRness of projects and related human processes by either external evaluators or the researchers themselves. We discuss better ways to implement simple FAIRness assessment in various communities and identify procedures and training that must be deployed and adapted to their practices and level of understanding. The next steps for the SHARC IG will be to elicitate the requirements for FAIR literacy in support of emerging practices for the FAIRification of data and services and how credit and reward mechanisms could advance data sharing.

Keywords

FAIR principles; FAIRness literacy; FAIR assessment, Research data sharing; Research evaluation; FAIRification; Pre-FAIRification; rewarding; crediting.

Introduction

The importance of sharing research data for scientific progress is becoming more and more accepted and increasingly required by funders. Still it is very challenging. The lack of recognition for the work required (i.e., preparing data for sharing that will allow effective reuse while recognizing appropriate attribution) is among the obstacles (Curty *et al*. 2017). The SHAring Rewards and Credit Interest Group (SHARC IG) was established as part of the Research Data Alliance (RDA) to analyze and improve scientific crediting and rewarding mechanisms for scientists who strive to organise their data (and material resources) for community sharing. The Interest Group has operated from the end of 2017 to 2020 and this paper presents some of its outcomes.

As part of the international initiative FORCE11[1], the FAIR principles have been proposed to make data *findable, accessible, interoperable and reusable* (Wilkinson et al., 2016). Within the SHARC IG, these principles are the prism used for encouraging and evaluating data sharing. Indeed, making resources available for community means ensuring data (and related materials) are findable and accessible on the Web (Mabile *et al*. 2016), comply with shared and adopted international standards making them interoperable and reusable by others (Hansen *et al*. 2018, Lannom *et al*. 2019). Although it is a major concern for improving data sharing, the FAIR principles do not address per se data quality issues. They focus on mechanisms to facilitate and optimize data sharing which does not preclude one's responsibility in assessing the quality and appropriateness of the data one's reuses. Accordingly, in this paper, we focus on assessing FAIRness and improving FAIRness literacy, not on assessing data quality.

Increasing the understanding of FAIRness literacy (Box 1), as well as assessing FAIRness are critical steps to make data sharing happen. In practice, from the researcher's point of view, making data FAIR requires considerable time, energy and expertise. Some technical solutions for making data FAIR are now well described (Wilkinson *et al.* 2017) and many groups (including within RDA) are working to develop methods and tools to transform the FAIR principles into practices. Infrastructures for preserving, locating and reusing research data already exist (Wittenburg *et al.*, 2019) such as the ones built in Europe in the context of the ESFRI roadmap[2] and the implementation of the European Open Science Cloud (EOSC).[3] In addition, worldwide validated certifications are now possible such as the ones delivered by the

[1] https://www.force11.org/
[2] https://www.esfri.eu/esfri-roadmap
[3] https://www.eosc-portal.eu

Core Trust Seal[4] which demands 16 trustworthy requirements for approved Core Trustworthy Data Repositories. Even for the most virtuous researchers producing data, the FAIR principles are usually not implemented rigorously and many recognize that taking into consideration every FAIR principle is not a priority.

On the funder's side, handling of research data according to the FAIR principles as part of Data Management Plans (DMPs) is now highly encouraged by most funders and even required in certain funding schemes. This holds true for many applications to European Commission calls including H2020 and for some national funding agencies (e.g., French ANR, [5] US NIH,[6] UK Wellcome Trust,[7] Austrian FWF[8]). However, projects that opt out are still encouraged to submit a FAIR compliant DMP on a voluntary basis (Jones *et al*. 2019). The Research Data alliance has several groups working on specifying DMPs, establish recommendations for their format and building mechanisms to facilitate their creation.

How to implement FAIR principles is still not well defined; concrete guidelines in terms of standards, tools, are needed, and developing progressively a deep FAIR culture is an essential preparatory step for FAIRification. One crucial step to foster FAIRification and to develop the FAIR culture is to reward compliance with FAIR principles as a first-class research output (Cambon-Thomsen *et al.* 2011). To that aim, appropriate human-readable criteria must first be clearly defined to enable FAIRness assessment and methods and processes must be developed to enable FAIRification.

Data should be made as FAIR as possible, even if it is not possible to make data open (Landi et al. 2019). However, with the new age of Open Science, the *European Commission Working Group on Rewards under Open Science*[9] delivered a frame matrix to help implementing and assessing FAIR criteria during the elaboration of DMPs (EC DGRI, 2017) .

Based on this matrix, Reymonet *et al.* (2018) made a detailed plan of quality criteria required during the three phases of development of DMPs (before, during, after). Wilkinson *et al.* (2018) published an article displaying the FAIR criteria according to their ability to be validated automatically. We believe these criteria are not easily put in practice by researchers and are difficult to interpret by evaluators. Therefore, many recent collaborative works propose ways to implement, adapt and evaluate FAIR principles in several communities (e.g., Herschel *et al.*

[4] https://www.coretrustseal.org

[5] https://anr.fr/en/latest-news/read/news/the-anr-introduces-a-data-management-plan-for-projects-funded-in-2019-onwards/

[6] https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

[7] https://wellcome.ac.uk/funding/guidance/how-complete-outputs-management-plan

[8] https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management/

[9] https://ec.europa.eu/research/openscience/index.cfm?pg=rewards_wg

2017, Doorn, 2018, Doorn & Timmermann, 2018, Federer *et al.* 2018, Mons *et al.*, 2017, Stall *et al.*, 2018, de Miranda Azevedo & Dumontier, 2019, Erdmann *et al.,* 2019, Sansone *et al.*, 2019). However, we are also reaching a moment where the FAIR principles now need cross-community convergence and *consensus* (EC DGRI, 2016; EC DGRI, 2018; Jacobsen *et al.*, 2019; Sustkova *et al.*, 2019; Thompson *et al.*, 2019; Wilkinson *et al.*, 2019). The work on FAIR data standards, repositories and policies is already ongoing as very well illustrated by the FAIRsharing.org platform which gathers more than 2800 registered standards, databases and policies (McQuilton *et al.*, 2019, Sansone *et al.*, 2019). Besides, one may mention other initiatives such as other RDA WG (*e.g.*, FAIR data maturity model) or the international GO FAIR initiative envisioned by Barend Mons and supported from the very beginning by the Netherlands, Germany and France. Implementation of the FAIR principles is a process that must be thought of in a progressive and community-oriented way. It must be consolidated within existing practices to ensure that they evolve without interruption and in a way that is acceptable to the various actors. However, in order to facilitate this gradual implementation and assessment by people in scientific communities, more work is needed.

This paper presents the design and assessment of a method developed on the basis of the previously mentioned related work to enable FAIRness assessment either by external data evaluators or by researchers themselves. It It also addresses the building of pre-FAIRification processes which are needed by the community to better understand the requirements and stakes in FAIRification. Our work is meant for all stakeholders in research data sharing processes.
This work results from workshops (Breakout sessions at RDA 13th and 14th plenaries), a survey and multiple discussions (teleconferences) that took place as part of the RDA-SHARC IG from 2017 to 2020.

Defining key FAIR criteria and designing an evaluation tool

Glossary

Because of the interdisciplinary nature of our IG work, various interpretations of underlying concepts have arisen implying a need for clear definitions. A glossary has been created using as many community-approved references as possible (Box 1), as well as other complementary terms in supplementary material section. Definitions that could not be referred to in the

scientific literature were all discussed and approved within the SHARC IG and validated by volunteers from the RDA community.

Community-approved: Validated by any specialized and recognized authority endorsed by the researcher's community.
Derived data: Data that is a result of substantial change like enrichment, modification or transformation of raw or primary data.
FAIRness: Level of FAIR criteria compliance.
FAIRification: process of technical implementation of the FAIR principles regarding data, i.e., a stage where the concepts that will be related to data and metadata are defined or chosen within existing ontologies, and then implemented. The results of FAIRification on FAIR data can be queried by a machine.
FAIRness literacy: Level of understanding and sharing of knowledge of the FAIR principles, their importance and their implementation acquired within a given disciplinary community. It can only be evaluated using criteria understandable by humans (i.e., a progression of skills beginning with the ability to understand the principles, and culminating in experiences to implement them).
FAIRness assessment: The use of criteria to assess/evaluate to which degree data respect the FAIR principles.
FAIR ontology: Definition and organisation of concepts that must be used to permit FAIR efficiency for data sharing. One can use the word ontology as an overarching term for any semantic resource or knowledge organization systems e.g., terminologies, thesaurus, vocabularies. Such resources are needed for FAIRification of data, but as any other data they also need to be FAIR.[10]
Human-readable: Property required for criteria that can be understood by any scientist including non information technology people
Metadata semantic validator: Tool to check validity of metadata obeying semantic rules.
Ontology: Vocabulary describing in a formalised machine readable language () the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms)
Pre-FAIRification: Processes required in a community to permit the FAIR principles implementation driven by a sufficient level of FAIRness literacy of each stakeholder. It needs in particular adapted processes of awareness raising, training and planning in existing project management initiatives.

Box 1: Glossary of terms used in the SHARC IG approach.

[10] https://www.slideshare.net/EUDAT/fair-data-requires-fair-ontologies-how-do-we-do

Key FAIR criteria

Encouraging implementation of efficient data sharing methods requires that they can be assessed easily qualitatively and quantitatively by any scientist. In particular, the FAIR data principles must be intelligible by any researcher, even if he/she is not a data scientist. FAIRness assessment must be realistic and pragmatic – what should be measured, how to explicitly find the information needed.

On the basis of the work done by Wilkinson *et al.* (2016; 2018) and Reymonet *et al.* (2018) in particular, the SHARC IG's work has resulted in a new classification of FAIR criteria according to the questions researchers should address while elaborating DMPs. An extensive set of criteria has been elaborated and organised as illustrated by the mind map Figures 1-4 and further developed in Tables 1-5.

FAIR criterion type	Short name	Proposed criteria	ID FAIR criterion
Sharing motivation	Nature of data shared	Beyond the ones with restricted access, are all relevant datasets shared or available for sharing over the evaluation period?	M01_E
	Existing DMPs	Does the researcher write appropriate / community-approved DMPs explaining their sharing strategy and methods? If so, specify if a tool/template has been used.	M02_E
	Long term preservation strategy	Is a long-term data preservation strategy properly provided within appropriate or community-approved infrastructures? (e.g., in a sustainable data archive)	M03_E
	Nature of sharing motivations	Which motivations are declared by the researcher?	M04_E
	Any data sharing training	Has any specific training for data sharing been followed by the researcher? If so, which programme.	M05_R
	Use of open community software platforms	If relevant, does the researcher use any approved or appropriate open community	M06_R

		software platform (for development, use and/or reports)? If so, which platform.	
	Existing SMPs	If relevant, does the researcher write Software Management Plans (SMPs) explaining all software used and their sharing strategy and methods for the software reusability? If so, specify which tool/template.	M07_D
	Other motivation	Any other motivation?	M08_D

Table 1. Detailed list of FAIRness assessment criteria related to 'Motivations'.

FAIR criterion type	Short name	Proposed criteria	ID FAIR criterion
Identification	Indexed identifier	Are each data/dataset identified by an indexed and independent identifier?	F01_E
	Unique, global, persistent identifier	Are the data identifiers unique, global and persistent?	F02_E
	Identifier scheme	Has any identifying schema been used for data (e.g., DOI)?	F03_E
Metadata traceability	Persistent metadata & data link	Are the metadata linked to the dataset through a persistent identifier?	F04_E
	Persistent metadata & authority link	Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?	F05_E
	Persistent datasets & authority link	Are all datasets linked to an authority over time (legal entity, e.g., institution, association or established body) through a unique and persistent identifier?	F06_E

Metadata description & searchability	Standards / dictionary for data description	Choose the question corresponding to your case: if standards exist, has the researcher used valid and updated standards for data description AND are the data standards approved by community- or appropriate authorities? If no standards exist, has the researcher created a well described community-approved data dictionary?	F07_E
	Search keywords	Does the researcher use community-approved well-described search keywords from a controlled vocabulary / thesaurus / ontology?	F08_R
	Metadata semantic validator	Does the researcher use a semantic validator for metadata (recommended by standardisation authorities or community-approved)?	F09_R
	Data format / type description	Are the types and formats of data generated / collected well described as recommended by the community or by standardisation authorities?	F10_E
	Versioning system	Does the researcher use a community-approved well described versioning system (for a proper tracking of versions)?	F11_R
	Metadata format validator	Does the researcher use any community-approved / appropriate metadata format validator recommended by standardisation authorities?	F12_R

Table 2. Detailed list of FAIRness assessment criteria related to 'FINDABLE'.

FAIR criterion type	Short name	Proposed criteria	ID FAIR criterion
Repository	Data storage in repositories	Does the researcher use data repositories for the storage of data?	A01_E
	Repository with quality criteria	Does the researcher choose a repository with quality criteria displayed by repository authorities?	A02_R
	Repository with guidelines for quality standards	Do the chosen repositories use guidelines for quality standards?	A03_D
Data security and services	Efficient and rich services for various uses & users	Does the researcher use efficient and rich services to access data (various formats, visualisations, practical tools and systems adapted to different types of use and users)?	A04_E
	Quality of description for integrity of data	Are well described and community-approved criteria / standards used for assessing the integrity of data (e.g., data veracity, data consistency during expected time / space / sampling)?	A05_R
	Quality of migration tools for data access	Regarding data availability, does the researcher make the necessary provisions as to the migration of tools for continuous and persistent access by the authorized users?	A06_R
	Quality of information for data reproducibility	If applicable, regarding data reproducibility (experimental and/or modeling data), does the researcher provide all the necessary information to ensure good reproducibility of data including the functioning of the tools, operating systems for continuous and persistent processes?	A07_R

	Quality of backup processes	Regarding data preservation, is the quality of backup and migration processes on the long term sufficient?	A08_R
Access restrictions	Data access restriction justification	In case of a non legal restricted access, is the restriction properly justified by the researcher?	A09_E
	Level of security for sensitive data access	Regarding confidentiality: does the researcher provide a satisfactory security level of access and preservation for "sensitive data" as defined by legislation or by the community?	A10_R
	Sharing beyond legal restrictions	Are the reasons for not sharing datasets always based on legal / regulatory rules?	A11_R

Table 3. Detailed list of FAIRness assessment criteria related to 'ACCESSIBLE'.

FAIR criterion type	Short name	Proposed criteria	ID FAIR criterion
Interoperable	Standard vocabularies, thesaurus, ontologies or data dictionary?	Are standard vocabularies, thesaurus or ontologies used for all data types present in datasets, to enable interdisciplinary interoperability between well-defined domains? If not, is a well-defined open data dictionary provided?	I01_E
	Description of interoperability criteria	Are the interoperability criteria explained?	I02_E
	Quality of data models	Are well-described and community-approved data models used?	I03_R
	Quality level of access for datasets description	Is there a high level and easy access description of the datasets available on the Web? (e.g., as a data paper, landing page)	I04_R

	File naming conventions	Have community approved well-described naming conventions been used for the files?	I05_R

Table 4. Detailed list of FAIRness assessment criteria related to 'INTEROPERABLE'

FAIR criterion	Short name	Proposed criteria	ID FAIR criterion
Data potential	Actions for data reuse potential	Which relevant actions have been undertaken by the researcher to enhance the data reuse potential?	R01_E
	Open format and open source software	Has the researcher favored open formats and open source software, particularly for purposes of reuse; if not, does he/she use standard or well documented and shared formats within the community? If the researcher uses proprietary format, does he/she propose tools to facilitate the translation of the data in open format?	R02_R
	Description of potential users and needs	Are the data/software potential users identified and their needs well described?	R03_R
	Description of potential reuse for each user	Is the data/software reuse potential identified and well described with defined objectives for specific types of users?	R04_R
	Evidence of reuse	Have the datasets been reused (evidenced by a citation, acknowledgement or else)?	R05_D
Data traceability	Provenance for raw and derived data	Are the provenance and type of all data properly specified (origin of raw, primary, derived, processed, derived and secondary data)?	R06_E
	Description of data processing steps	Is each step of data processing well described and traceable?	R07_R

Data enhancement	Description of data curation protocol	Has the researcher documented an intelligible, well-described community-approved data curation protocol including the curation journal that tracks changes for each dataset?	R08_R
	Description of data enrichment / requalification processes	Does the researcher have any reliable / community-approved data enrichment / requalification process to increase reusability with several standards / thesaurus / ontologies?	R09_R
	Description of data quality assurance processes	Are the data quality assurance processes well described and community-approved?	R10_D
Reusability tools	Description of methods and tools that permit long term integrity and understandability of data	Does the researcher provide information on methods and tools that permit the understandability, integrity, value and readability of data intended to be kept on the long term? (e.g. versioning, archival and long term reuse issue for protocols, softwares, required methods and contexts to create, read and understand data)	R11_E
	Description of time range	If relevant, is the time range for which the data remain re-usable specified?	R12_R
	Data operating tools and their documentation	If relevant, are all data operating tools and the associated documentation needed to access and reuse the data included? (methods, instruments measurement, software, survey, analysis, observation, compilation, simulation, etc.).	R13_R
	Quality standards for enriched data	Does the researcher provide quality standards where appropriate on enriched data for reuse?	R14_D

Reusability right	Data sharing arrangements meeting data ethics and protection requirements	Do the data reuse control and data sharing arrangements meet the data protection and local/national ethics requirements?	R15_E
	Justification of legal reuse restriction	In case of a legal reuse restriction (such as personal data, state and public security, national defense secret, confidentiality of external relations, information systems security, secrets in industrial and commercial matters), is the restriction properly justified?	R16_E
	Quality of licenses / rules for large reuse	When possible, does the researcher specify community-approved licenses or rules (free, shared and well documented) that allow a large data reuse?	R17_R

Table 5. Detailed list of FAIRness assessment criteria with regards to 'REUSABLE'.

The assessment tool

On the basis of the defined criteria (Tables 1-5), an assessment tool has been designed as a chart for the evaluator to assess the quality of the researcher/scientist sharing practice, over a given period, taking into account the means and support available over that period. This tool can also be used as a self-assessment tool conceived as a checklist for scientists to identify if their own practices are compliant with the FAIR principles and to pinpoint the hurdles that may hinder their efficient sharing and reuse of data.

This tool is built to be as generic as possible and as discipline-specific as needed. It is designed to be as much as possible intuitive (i.e., understandable for non-IT people, especially for interdisciplinary networks), but specific enough to measure the degree of FAIRness compliance. In addition, it aims to be easy to complete, requires information that reflects the real work of scientists, and takes a step-by-step approach to validate different stages of FAIRness.

The tool uses a mind-mapped tree-graph structure based on previous works on FAIR data management. It displays inter-relations that have been analyzed (*e.g.,* 'if something does not comply with this criterion it cannot fulfill the related one').

The tool relies on about 60 key criteria organized and categorized as shown in Figures 1-4 and Tables 1-5. For each criterion or question, four assessment levels have been proposed:

‘Never or Not assessable / Mandatory / Sometimes / Always. One degree and only one needs to be selected while assessing.
The three levels of priority identified for each criterion (essential / recommended / desirable) and the two types of assessment grids, one extensive with all criteria (over 60 criteria) and one ‘simplified’ grid with essential ones only (18 criteria) were developed to enable a gradual incrementation of scientists’ efforts.
The simplified criteria table does not permit to assess FAIRness rigorously, but allows a first assessment by any evaluator based on all essential criteria that are required to be FAIR compliant.
The assessment of each criterion enables us to compute a final score for each criterion that is calculated from the selected responses in comparison with the total number of criteria in each group. The motivations for sharing are incorporated qualitatively in the final interpretation (Figure 5). It is likely that for one‘s first FAIRness assessment (initial state of FAIRness), an important part of the criteria will never be fulfilled (often when people do not understand the criteria or do not know how to apply them). When ‘Mandatory’ is checked, that means that they know how to do it, but all kinds of means dedicated to FAIRness are minimum. ‘Sometimes’ is checked when stakeholders know that FAIR compliance is important, but they can not enforce it every time (due for instance to time/resource/skills limitations). ‘Always’ is the case that means optimal organisation (as infrastructure design or as social consensus) but where quality processes should be maintained for long term compliance (template in the supplementary material; evaluator instructions in the ‘*Read me*’ tab).
https://docs.google.com/spreadsheets/d/1vloqbekIGlqiDwzE9jqZzoaoDCbwYQlxOWbZzIxIYbI/edit#gid=448406479

Assessing the tool

Getting intelligible, realistic and human-readable assessment criteria would help to guide scientists to follow the FAIR data principles and would help evaluators to objectively achieve their task. To that aim, an online assessment survey was designed within the SHARC community to seek feedback from volunteers on clarity and usability of the criteria and tools proposed (supplementary material).
We exposed the two assessment tools (extensive and simplified) to volunteers from the SHARC IG and a panel of external researchers active in data management at breakout sessions of 13th and 14th plenaries. We asked them to critique the tools for their completeness, relevance, efficiency, clarity, priority levels and the definitions. The comments and suggestions arising were compiled and analyzed.

In particular, the need for a progressive implementation of the FAIR criteria quickly emerged, where at each stage the levels of achievement rewarded would be made explicit.

Need for a gradual implementation of FAIR criteria

Using the collective feedback from the survey – assuming it might not be fully representative – we formalised the steps required for FAIR data sharing as a step-by-step iterative process as described in parts A to C and further discussed in parts D to F.

A) Fostering the pre-FAIRification decision

Observation:
A better perception of the gain and return on investment from the use of controlled vocabularies (especially in the form of ontologies) triggers the need to prepare FAIRification.
Discussion:
The following discussion came out from the collective feedback and provided arguments to motivate researchers and institutions to invest sufficient means (human or financial) for the needs of structured and long term FAIRification processes.

Each of the four categories of FAIR criteria refer to the use of community-approved formats, standards and controlled vocabularies (Kryger Hansen *et al.* 2018). Common naming conventions for classifying standards for reporting and sharing data or other objects appear favourable, assuming that those standards are registered and therefore reusable and mappable. The use or implementation of controlled vocabularies or ontologies may seem complex and time-consuming, especially since the means to find and explore them are still not fully harmonized (i.e., in the biomedical domain there are multiple well known and used terminology services or ontology repositories that are not systematically displaying the same content, or offer the same APIs). The responsibilities lie, among others, with repositories and standard providers. They have to outreach into the community via targeted education, publicity at events and reporting success stories. However, the first level of interoperability can be easily reached, with substantial return on investment by using vocabularies recommended by the W3C to make data of different domains and disciplines compatible with each other. Semantic Web standards such as RDF, OWL and SKOS recommendations from the W3C, help metadata machine-readability.

The use of these reference vocabularies or ontologies also improves indexing and findability of data and sometimes makes it possible to cite an author for a selection of data only. This can help giving additional impact to original studies by generating more citations. Standard

vocabularies and ontologies also allow inclusion in many catalogues formatted with these recommendations. It also offers a greater ability to maintain data on the long term, allowing much greater backward compatibility of new software and new innovative processing chains of old data. Further, the use of vocabularies and ontologies increases the number of links and therefore possibilities of access and reuse of all or parts of a dataset. One can also assume that in the future a major criterion of evaluation of DMPs will be building data with community shared vocabularies and ontologies.

B) Planning pre-FAIRification training and support

Observation:
The survey results show that essential criteria are not always understandable without specific education or training. Often, the respondents estimate that implementation of some criteria can be time consuming and may need some technical support.
Discussion:
The adoption of the FAIR data principles requires a cultural change. As a need for pre-FAIRification process, the first desirable and desired step for researchers is to better understand the efforts required to meet each FAIR criterion in a suitable manner that optimizes costs (time, human resources, skills). This good comprehension of the necessary efforts is a *sine qua non* condition for the researcher. These efforts can relate to:

- Policies: for example, the choice of the perimeter of the community in which this FAIRification process will be implemented;
- Strategic and tactical steps: technological choices, and implementation steps for wide acceptance in the predefined community;
- Pedagogical steps: for example providing educational kits adapted to different levels of skills up to “qualified training” where trainees knowledge is evaluated;
- Human resources: enabling researchers, engineers and technicians to acquire different time scales, the necessary skills;
- Governance: the training of managers to improve the relevance of their arbitrations and improve the planning of the means necessary for FAIRification in the short and long term.

It is also necessary that all the benefits of the FAIRification are clearly explained and demonstrated. Beyond the constraints of FAIRification recently imposed in many calls for projects lies the risk to lose good scientific projects if the FAIR data principles are not (yet) well understood or not (yet) properly supported. Among the benefits of FAIRification is the impact of research, where data is a new entry point. It is also time saving for setting up new data processing chain, manipulating data, cleaning up data, reconstructing missing data, or feeding more rigorously machine learning processes. FAIRification of data or services is also expected

to build Virtual Research Environments (VRE) that are more reliable and allow for more reproducible research.

During planning (the final step in pre-FAIRification), the strategic and tactical choices for implementing FAIRification must allow for a gradual implementation. Each step must be understood by each of the stakeholders, the decision process about criteria prioritisation should be defined (this may depend on the area of application) and possibly the list of criteria that can be neglected if the means are not sufficient. For each criteria, it will be useful to show the return on investment, especially in terms of scope, speed, quality and richness of shared treatments.

C) Planning a step-by-step pre-FAIRification process

Observation:
Our discussions while elaborating this list of criteria, revealed the following elements: considering the diversity of actors whose contribution is necessary for FAIRification, it is imperative to rely on planning tools and on structuring the organization of the actors and on prioritizing the steps required with respect to each one of these actors, depending on the means and skills available.
Discussion:
Pre-FAIRification is a *sine qua non* step for community acceptance of the efforts required for later FAIRification; especially with an efficient long-term effect and considering interoperability between heterogeneous systems.
Pre-FAIRification involves different stakeholders, such as funders, policy makers, publishers (to make FAIR data a requirement), institutions (to provide infrastructure, training, support and policies at their departments e.g., in the library), and disciplinary communities (to create community standards). This illustrates that researchers have to be supported throughout the entire life cycle of research data starting from generation until final sharing and archival.
To empower scientists for such a new work, it is necessary to co-construct the FAIRification planning tools by taking into account the resources of each actor (e.g., typically DMPs in a research project are under the responsibility of the project coordinator). One can expect future research work-programmes project (e.g., European Commission's Horizon Europe will be granted only if the FAIRification processes are properly detailed and correctly sized in the proposal (human and technical resources in particular).

Identification of community milestones to organise FAIRification as an iterative and quality process is a result of our discussions (Figure 6). FAIRification includes four distinct steps in an iterative process:

1) Preparing: FAIRification for a specific scientific community requires first that what is meant by FAIRification is carefully explained and that constraints and advantages in the short, medium and long term are described.
2) Training: to improve FAIRness literacy; it permits to convince stakeholders in the whole community. Note that this literacy should be maintained during the planning for FAIRification, especially when this is done before funding.
3) Pre-FAIRifying: this stage must be feasible for all actors. It encompasses the largest common denominator of objectives achieved by all and its success is crucial to empower the whole community further in the FAIRification process.
    a) Defining the community: a set of actors agreeing on the way they delimit their own community and the subjects related to it.
    b) Defining objects and variables to be FAIRified
    c) Selecting what to identify and index (data, actors, objects, processes, etc.)
    d) Analyzing which are the common denominators
    e) Reducing the explicit needs and expectations related to this first step in order to ensure the achievement of a common objective for all stakeholders (downward levelling)
4) FAIRifying by applying the prepared plan, check the results are compliant with community approved plan, and adjust if necessary.

Whenever a common goal is reached, the community can redefine a new scope (and ideally enlarge it). Planning the second iteration with a new statement of expectations (return to step 3b or 3a if the community has evolved, which is very often the case) requires before so “Preparing” and “Training” steps.

D) Dealing with the unequal understanding of criteria in the pre-FAIRification process

Observation:

As often, and even within a particular scientific community, definitions and interpretations of concepts covered by the FAIR evaluation criteria are manifold. The diversity of perception of minimum work to be achieved in the short and long term often leads to either a deep underestimation of the means to be implemented or even worse, to the fact that when significant resources are allocated to FAIRification, the solutions provided do not meet the prerequisites and create additional locks.

Discussion:

Two impediments relate to a specific type of data communities:

- Regarding the more mature scientific communities, there may be a reluctance to change their tools and entrenched habits. FAIRification will probably require a great deal of effort but how quickly can we hope to reach a consensus?

- Concerning emerging sciences: communities are often embryonic; the complexity of the topics covered and the emergence of jargons (particularly in interdisciplinary fields) increase the difficulty of organizing complex processes around data at the interfaces of different scientific cultures.

These two impediments can be removed by detailing concrete actions and associated means for pre-FAIRification. Such actions must explain the meaning and the interest of the respect of each criterion within the framework of the community . In addition, these actions will not only have to be approved by the domain community as bringing a real improvement in the FAIR quality of data, but also meet all criteria for sustainable reuse beyond the community.
It is not entirely clear for each stakeholder whether or not criteria of FAIR assessment can be considered as part of quality data assessment or only data management assessment.
Moreover, increasing global FAIRness compliance requires absolutely to raise the level of FAIRness literacy for each actor, and particularly that of managers, who have the role of arbitrating between several objectives. For instance, other recruitments like statistician or manager assistant can be prioritized before recruitment and/or training of FAIRness skills.

E) Taking care of Sustainability

Observation:
Problem of long term maintenance, means and skills availability during long time and technology support over time.
Discussion:
The complex concepts behind FAIRification evolves quickly: Because of constant knowledge and process improvement, information systems for research data are fed by constantly renewed protocols. FAIRifying new and old data requires different iterative processes based on a consultation of all participants each time a new concept is needed. Both situations demand some differentiated human, technical or financial resources and organisations.
New data produced are increasingly heterogeneous and multi-source, sometimes even in exotic formats. Each new set of data, resulting from an exploration of a new research subject and / or the implementation of an experimental process based on emerging technologies is still rarely based on a fixed scheme allowing an immediate match with the FAIR principles. Human and technical means are critical for implementing them.
For old data, unless it is a research subject on its own, FAIRification can be funded currently only in response to calls for research projects re-using these data.
Whether these data are old or recent, FAIRification is only possible if it occupies a central place in the research project, and therefore is understood and adhered to by all stakeholders.
It is clear, even today in the call for projects on FAIRification, that the level of understanding of the FAIR issues and of the needs for FAIRifying is still very uneven.

F) Increasing good research data sharing during pre-FAIRification processes by rewarding and crediting

Observation:
Among the obstacles to data sharing is the lack of recognition. This is even more true when considering the efforts required for FAIRification which is critical for efficient data sharing.
Discussion:
Until now, research institutions do not take into account FAIR principles implementation when evaluating researchers. To that aim, it is essential that the FAIRification activity be assessed in research evaluation schemes (policy issue at institutional, national and supranational levels). Specific mechanisms necessary for the evaluation scheme (*e.g.,* identifying researchers and their data for rewards and credits) have to be included in the early FAIRification process. Those are first aimed to link shared data within the global web of data by associating their creators/contributors (with the help of metadata) so that shared data paternity is unequivocally attributed. Therefore we need perennial and unambiguous identification of data on one side and persons (creators, contributors) on the other side. Various types of identifiers currently exist, for research outputs such as DOI, ARKs, Handles, URIs etc., or for researchers (ORCID, ResearcherID, Scopus IDs, etc.) and for research organisations (ROR IDs, GRID, ISNI). With CrossRef[11] part of it, the current ecosystem makes it possible to interconnect unambiguously scientists with their deposited datasets, publications, other professional contributions or activities (e.g., open peer reviews), research organisations, funders and so forth. Online generated metrics can then help crediting the work and some kind of reward may follow if this is taken into account in the research evaluation scheme (project awarding; dedicated financial support; career promotion).

Conclusions

The RDA-SHARC Interest Group (IG) has established that the FAIR data principles need to be adequately explained from the very beginning of a research project design and that training needs to be provided as early as possible. To help implement this, the SHARC IG has developed a tool to support the assessment of FAIRness literacy and therefore enable measurement of progress towards compliance.

A step-by-step process will help teams organise the various actions towards the achievement of FAIR data, (i) by pre-FAIRifying (fostering decision; planning training & support; planning step-by-step process), and (ii) by anticipating heterogeneity in FAIR literacy and sustainability.

[11] https://www.crossref.org

By defining a simple first step, prioritizing the criteria to be implemented, is a way of “learning to walk”. Even if the FAIR goal is not reached in one step, it can make FAIRification more understandable and improve its acceptability (e.g., by using templates and/or data papers).

We highlight that researchers should be supported by data management professionals (not only data stewards), organised in networks, and embedded in institutions. In order to enhance treatment of data according to the FAIR principles, we suggest that organisations should be assessed on the basis of how well they support their researchers in becoming FAIR advocates.

Perspectives

At present, research communities lack pragmatic examples of the added value of FAIRification, especially since the implementation of the FAIR principles and the criteria that depend on them. These (i) are not related to a particular technology, (ii) are strongly dependent on communities and their themes, and (iii) concern acquisition systems whose continual mutations are an intrinsic characteristic. FAIRification needs to be accomplished in stages adapted to each community. For example, the recommendation level for a given criterion (essential / recommended / desirable) may vary from one research community to another (e.g., criterion R17_R or R17_E).

A prerequisite for the implementation of the majority of the criteria presented in our assessment tool is the adoption of a consensus on methods and timelines for implementation. Achieving this consensus, however, requires first developing a sufficient level of FAIRness literacy within each professional community, and in each member organization of the community. Special emphasis should be placed on highlighting the differences in understanding the human-readable FAIR compliance criteria between various users (researchers and evaluators, young and senior researchers, IT professionals, etc.) within particular communities.

Several other possible open perspectives arise from the SHARC IG work: (i) building a gradual assessment of FAIRness literacy for researchers; (ii) helping identify needs to build FAIRness assessment guidelines for better sharing capacity by researchers (based on rewards and credit); (iii) developing step-by-step curation processes, technical assistance, and training for FAIRness compliance implementation; (iv) inviting each community to test methods and tools, adapt them and develop their own adapted assessment tool; (v) creating and installing Good FAIRification Practice; (vi) deploying efforts to put in place a standard specifically dedicated to FAIRification process; and (vii) improving and implementing FAIRification processes in response to case study/stakeholders’ selection in order to avoid a biased assessment.

FAIR compliance will certainly not be enough to enhance real and high data sharing and reuse; other mechanisms and qualities should be considered in future sharing processes (e.g., data veracity, quality, publicity, large indexation, curation, support). Developing a strategy to orchestrate efforts across the variety of communities seems to be a first priority in order to avoid the dispersed attempts of standards adoption and of compliance to FAIR.

ACKNOWLEDGEMENT

This work was partly supported by the Research Data Alliance, especially the "RDA Europe 4.0" project (H2020 grant N°777388), the EPPN2020 project (H2020 grant N°731013), the 'Infrastructure Biologie Sante' PHENOME-EMPHASIS project funded by the French National Research Agency (ANR-11-INBS-0012) and the 'Programme d'Investissements d'Avenir'.
This research is also a product of the PARSEC group funded by the Belmont Forum as part of its Collaborative Research Action on Science-Driven e-Infrastructures Innovation (SEI2018) and by the French National Research Agency (ANR-18-BELM-0002-02).
Complementary support was provided through the FAIRplus project that has also received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 802750. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA".
This manuscript publication is supported by the RDA Europe 4.0 project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777388.
Special acknowledgement to Julien Lecubin from OSU Pytheas, France for technical support in the survey and to all the participants. Laurent Dollé is grateful to project support by Innoviris.

COMPETING INTERESTS:

The authors have no competing interests to declare.

REFERENCES

Cambon-Thomsen, A., Thorisson, G.A., Mabile, L., *et al*. BRIF Workshop Group, 2011. The role of a bioresource research impact factor as an incentive to share human bioresources. *Nat. Genet.* 43, 503–504. https://doi.org/10.1038/ng.831

Curty, R.G., Crowston, K., Specht, A., *et al*., 2017. Attitudes and norms affecting scientists' data reuse. *PLOS ONE* 12, e0189288. https://doi.org/10.1371/journal.pone.0189288

de Miranda Azevedo, R., Dumontier, M., 2019. Considerations for the Conduction and Interpretation of FAIRness Evaluations. *Data Intelligence* 285–292. https://doi.org/10.1162/dint_a_00051

Doorn, P., Science Europe, 2018. Science Europe Guidance. Presenting a Framework for Discipline-specific Research Data Management. [*WWW Document*]. URL http://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pd

f [Accessed January 07, 2020]
Doorn, P., & Timmermann, M., 2018. Towards Domain Protocols for Research Data Management (IG Domain Repositories RDA 9th Plenary meeting Community-driven Research Data Management). Paper presented at the 9. *Plenary meeting Community-driven Research Data Management*, Barcelona. https://www.rd-alliance.org/sites/default/files/attachment/RDA%20DRIG%20Domain%20Protocols%20V3%20Barcelona%20April%202017%20-%20DoornAerts.pptx [Accessed January 07, 2020]
Erdmann, C. ,Simons N., Otsuji R., *et al*., 2019. Top 10 FAIR Data & Software Things. [*WWW Document*]. https://doi.org/10.5281/zenodo.2555498 [Accessed January 07, 2020]
Federer, L.M., Belter, C.W., Joubert, D.J., *et al*., 2018. Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS ONE* 13, e0194768. https://doi.org/10.1371/journal.pone.0194768
Gomez-Diaz, T., Romier, G., 2018. Research Software Management Plan template V3.2, PRESOFT project. [*WWW Document*]. URL http://www.france-grilles.fr/wp-content/uploads/2018/04/ModeleSMP_PRESOFTV3.2.pdf [Accessed January 07, 2020]
Hansen, K. K., Buss M., Sztuk Haahr L., 2018. A FAIRy tale. Zenodo. [*WWW Document*]. URL https://doi.org/10.5281/zenodo.2248200
Herschel, M., Diestelkämper, R., Ben Lahmar, H., 2017. A survey on provenance: What for? What form? What from? *The VLDB Journal* 26, 881–906. https://doi.org/10.1007/s00778-017-0486-1
Jacobsen, A., de Miranda Azevedo, R., Juty, N., *et al*., 2019. FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence* 10–29. https://doi.org/10.1162/dint_r_00024
Jones, S., Pergl, R., Hooft, R., *et al*., 2019. Data Management Planning: How Requirements and Solutions are Beginning to Converge. *Data Intelligence* 208–219. https://doi.org/10.1162/dint_a_00043
Landi, A., Thompson, M., Giannuzzi, V., *et al*., 2019. The "A" of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence* 47–55. https://doi.org/10.1162/dint_a_00027
Lannom, L., Koureas, D., Hardisty, A.R., 2019. FAIR Data and Services in Biodiversity Science and Geoscience. *Data Intelligence* 122–130. https://doi.org/10.1162/dint_a_00034
Mabile, L., De Castro, P., Bravo, E., *et al*., 2016. Towards new tools for bioresource use and sharing. *Information Services & Use* 36, 133–146. https://doi.org/10.3233/ISU-160811
McQuilton, P., Batista, D., Beyan, O., *et al*., 2019. Helping the Consumers and Producers of Standards, Repositories and Policies to Enable FAIR Data. *Data Intelligence* 151–157. https://doi.org/10.1162/dint_a_00037
Mons, B., Neylon, C., Velterop, J., *et al*., 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37, 49–56. https://doi.org/10.3233/ISU-170824
Reymonet, N., Moysan, M., Cartier, A., *et al*., 2018. Réaliser un plan de gestion de données « FAIR » : modèle. [*WWW Document*]. URL https://archivesic.ccsd.cnrs.fr/sic_01690547/document [Accessed January 10, 2020]
Sansone, S.-A., McQuilton, P., Rocca-Serra, P., *et al*., 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol* 37, 358–367. https://doi.org/10.1038/s41587-019-0080-8
Stall, S., Cruse, P., Cousijn, H., *et al*., 2018. Data Sharing and Citations: New Author Guidelines Promoting Open and FAIR Data in the Earth, Space, and Environmental Sciences. *Science Editor* 41, 83–87. https://www.csescienceeditor.org/wp-content/uploads/2018/11/CSEv41n3_text_83-87.pdf [Accessed January 07, 2020]
Sustkova, H.P., Hettne, K.M., Wittenburg, P., *et al*., 2019. FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources. *Data Intelligence* 158–170. https://doi.org/10.1162/dint_a_00038
Thompson, M., Burger, K., Kaliyaperumal, R., *et al*., 2019. Making FAIR Easy with FAIR Tools: From Creolization to Convergence. *Data Intelligence* 87–95. https://doi.org/10.1162/dint_a_00031

European Commission Directorate General for Research and Innovation, 2018. Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data. Luxembourg: Publications Office of the EU. [*WWW Document*], URL https://op.europa.eu:443/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF [Accessed January 07, 2020].

European Commission Directorate General for Research and Innovation, 2017. Evaluation of Research Careers fully acknowledging Open Science Practices; Rewards, incentives and/or recognition for researchers practicing Open Science. Luxembourg: Publications Office of the EU. [*WWW Document*], URL https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf [Accessed January 07, 2020]

European Commission Directorate General for Research and Innovation., 2016. E.U. H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, Version 3.0. Luxembourg: Publications Office of the EU. [*WWW Document*], URL https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [Accessed January 07, 2020]

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., *et al*., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., Dumontier, M., Sansone, S.-A., *et al*., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data* 6, 1–12. https://doi.org/10.1038/s41597-019-0184-5

Wilkinson, M.D., Sansone, S.-A., Schultes, *et al*., 2018. A design framework and exemplar metrics for FAIRness. *Scientific Data* 5. https://doi.org/10.1038/sdata.2018.118

Wilkinson, M.D., Verborgh, R., Santos, L.O.B. da S., *et al*., 2017. Interoperability and FAIRness through a novel combination of Web technologies (No. e2522v2). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.2522v2

Wittenburg, P., Sustkova, H.P., Montesanti, A., *et al*., 2019. The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data [*WWW Document*], URL https://arxiv.org/abs/1902.11162? [Accessed January 07, 2020].

FIGURES

Figure 1. Mind mapping of FAIRness assessment criteria. The mind map shows the extensive set of criteria for the FINDABLE criteria (12 criteria). The level of importance of each criteria is divided into three categories (illustrated by three colours), Essential (purple) / Recommended (brown) / Desirable (red).

Figure 2. Mind mapping of FAIRness assessment criteria. The mind map shows the extensive set of criteria for the Accessible criteria (11 criteria). The level of importance of each criteria is divided into three categories (illustrated by three colours), Essential (purple) / Recommended (brown) / Desirable (red).

Figure 3. Mind mapping of FAIRness assessment criteria. The mind map shows the extensive set of criteria for the Interoperable criteria (5 criteria). The level of importance of each criteria is divided into three categories (illustrated by three colours), Essential (purple) / Recommended (brown) / Desirable (red).

Figure 4. Mind mapping of FAIRness assessment criteria. The mind map shows the extensive set of criteria for the Reusable criteria (17 criteria). The level of importance of each criteria is

divided into three categories (illustrated by three colours), Essential (purple) / Recommended (brown) / Desirable (red).

Figure 5. Example of the assessment template for the “ACCESSIBLE” criteria. In this example, A06_R, A08_R and A10_R are fulfilled only when they are required by grant providers, A01_E, A02_R, A04_E, A05_R, A11_R are sometimes fulfilled while it is not required, A03_D, A07_R are never fulfilled and A09_E is always fulfilled. in the case of a criterion in the “ACCESSIBLE” group is never fulfilled: if the criterion is not applicable, it will not be taken into account in the final evaluation. But if it is, the result of the assessment will recommend a strategy for a possible easy progress in FAIRification (community-approved choices and processes development).

Figure 6: FAIRification can be schematized as a wheel describing iterative quality steps that needs to be approved by the community throughout the process. This schema displays the “preparing” and “training” phases as *sine qua non* conditions of pre-FAIRification. The pre-FAIRification processes must be community-approved at each iteration. The FAIRification steps ‘check’ and ‘adjust’ implementation must be approved by the community before a new iteration.

Supplementary material:

FAIR assessment template; evaluator instructions in the ‘read me’ part.

https://docs.google.com/spreadsheets/d/1vloqbekIGlqiDwzE9jqZzoaoDCbwYQlxOWbZzIxIYbI/edit#gid=448406479