Seven Recommendations for Implementation of FAIR Practice

By the FAIR in Practice Task Force

Of the European Open Science Cloud FAIR Working Group

2020

Authors (in alphabetical order):

- Oya Deniz Beyan (0000-0001-7611-3501)
- Neil Chue Hong (0000-0002-8876-7606)
- Stefano Cozzini (0000-0001-6049-5242), AreaSciencePark,Trieste, Italy
- Marta Hoffman-Sommer (0000-0002-1451-5967)
- Rob Hooft (0000-0001-6825-9439)
- Liisi Lembinen (0000-0002-5176-2641)
- Juuso Marttila (0000-0003-3394-0341), University of Jyväskylä, Finland
- Marta Teperek (0000-0001-8520-5598), TU Delft, the Netherlands

With contributions from:

- Françoise Genova (0000-0002-6318-5028)
- Laura Molloy (0000-0002-5214-4466)
- Angus Whyte (0000-0002-5198-0833)

DOI: 10.5281/zenodo.3904140

This page is intentionally left blank. An executive summary will be added here in due time.

Introduction

The FAIR Practice Task Force was set up as one of the four task forces of the European Open Science Cloud FAIR Working Group. Its goal was to support the Working Group with an oversight of FAIR practices: learning about the possibilities of future FAIR implementation from current experience.

Even though the Task Force was not assigned a deliverable, this report was written because the Task Group's research into FAIR practices provided useful insights on gaps, differences and commonalities between communities. We wrote this report to share our findings and highlight the risks of not addressing these gaps.

This report can be seen as a followup on the 2018 report "Turning FAIR into reality" from the European Commission Expert Group on FAIR Data[1]. Our primary aim was to translate our findings into actionable recommendations to the decision making entities of the European Open Science Cloud (EOSC), as well as research funders and policymakers on how to turn FAIR into practice.

After a section describing our methodology and the limitations of our study, this report contains a disciplinary and a regional perspective on FAIR implementation. The disciplinary perspective summarises what we have been reading on fair practices split into 2x2 parts: *technical and social impediments* on one side, and *technical solutions and social enablers* on the other. The regional perspective shows trends in regional policies and how they have so far driven the development of FAIR practices, highlighting the differences and commonalities.

A separate section details FAIR practices for digital objects other than research data.

We close the report with two sections with insights. The first one describes where differences between disciplinary and regional implementations come from, what implications these differences have for policymakers turning FAIR into reality and how these differences can and should be addressed. Finally, we close off with our recommendations for the EOSC, research funders and policymakers.

This report was written collaboratively in an interesting time, with all authors working from home in the time of the Covid-19 pandemic in 2020. Our observations of data handling in this time helped us reflect that existing FAIR practices are already paying off for the expedited research processes needed to fight this new disease, but also that more acceleration would have been possible if FAIR practices would already have been implemented more broadly. There is still a lot to gain.

[1] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

Methodology

This section describes the methodology undertaken by the FAIR Practice Task Force of the FAIR Working Group in investigating FAIR practices, producing the body of knowledge document, writing this report and community consultation.

The investigation into FAIR practices was started with literature research (lasting from July 2019 to June 2020). Literature was organised as a body of knowledge in a dedicated online spreadsheet[2], to which various team members contributed reading resources. Reading resources were arranged by academic discipline[3]. The spreadsheet was open for community consultation and additional reading resources were contributed by various experts outside of the task force[4].

Each reading resource was allocated to a team member who analysed it in detail. Key findings from each resource were then classified into four different types: technical solutions, social enablers, social impediments and technical impediments. The applicability of each finding was then further matched to individual FAIR principles. Filtering was applied on the types and applicability of the different findings to facilitate easy, interactive queries of the spreadsheet content.

On 16-18 June 2020, the FAIR Practice Task Force met online to summarise their findings in a written report with recommendations primarily intended for the EOSC, research funders and policymakers. The members of the Task Force have different disciplinary backgrounds, which allows putting the different findings from the reading resources into perspective of FAIR data experts from the various fields, improving consistency. This written report thereby has become the symbiosis of conclusions from the reading list and members' own experience.

Stakeholder definitions used in this report are consistent with the terminology used in "Turning FAIR into reality" report[5], with the exception that a new stakeholder "EOSC" has been introduced defined as "those in decision-making capacity within the EOSC".

The report, as well as the body of knowledge spreadsheet, will be open for public consultation on 9 July 2020 (a dedicated webinar[6] will launch the consultation period). The consultation will close on 31 August 2020. Subsequently, both resources will be revised and finalised.

---

[2] Hooft, Rob; Beyan, Oya; Chue Hong, Neil; Cozzini, Stefano; Hoffman-Sommer, Marta; Lembinen, Liisi; … Teperek, Marta. (2020). FAIR in practice reference list (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3898674

[3] Outline of academic disciplines: https://en.wikipedia.org/wiki/Outline_of_academic_disciplines

[4] How to move from FAIR principles to FAIR practice? https://eoscsecretariat.eu/news-opinion/how-move-fair-principles-fair-practice - blog post announcing the work of FAIR Practice Task Force and requesting community contributions

[5] Turning FAIR into reality: https://op.europa.eu/s/n8cm

[6] WEBINAR: How to move from FAIR principles to FAIR practice? Current practices and recommendations for the future: https://www.eoscsecretariat.eu/events/webinar-fair-principles-fair-practice-recommendations-future

Limitations

This study has some limitations.

1. The body of knowledge was composed of reading resources known to FAIR Practice Task Force members, or recommended to them by external experts. Thus, the list of resources should not be perceived as an exhaustive information on fair practices.

2. Lack of data, or lack of information on practices should also be considered informative. Communities or subcommunities that are not aware of FAIR practices might be less likely to write publications analysing such practices, and also less likely to participate in surveys and research looking at FAIR practices.

3. The classification of findings by type and applicability was done as best-effort by the team member going through the resource, looking for the closest match. Therefore, there might be cases where a certain finding is classified as one type/applicability, but in fact could fit into more than one category.

4. Information on community practices is almost exclusively based on desk research and thus might not always be accurate, as it is based on (sometimes subjective) interpretations of the written text. In addition, attempts to engage with certain communities to verify information on their practices or get information about their practices were not always successful and/or are ongoing.

FAIR practices: a disciplinary perspective

An overwhelming majority of scientific references to the FAIR principles come from life and natural sciences[7]. Nevertheless, sufficient information is available about the practical implementation of FAIR practices across disciplines to make a general overview of what has been done already, and also to identify what stands in the way of a further deployment of FAIR within communities, both from technical as well as social perspectives.

Our observation is that, although the scientific needs differ between disciplines, and thus each discipline searches for its own solutions and follows its own path towards FAIR data, the difficulties as well as enablers encountered are often shared.[8]

Technical impediments

There are many generic and many data-type specific repositories. Nevertheless, some fields feel a **lack of specific repositories** (e.g. earth sciences) or **lack of repositories that can deal with complex outputs** (humanities) or **insufficient infrastructure for transporting and archiving of large data** to/from repositories. Also reported is a **lack of sufficiently flexible infrastructure for archiving of sensitive data**. On the other hand, we also encountered the complaint that there are **too many different repositories** to search for data.

Interoperability principles are widely considered the hardest to adopt. It is sometimes observed that efforts to improve FAIRness tend to be more focused on findability instead of interoperability, because this is easier to start with. Regarding intra-disciplinary interoperability we see that it is hard to make **traditional text-based outputs like lexicons and bibliographies** FAIR. On the other hand, some communities choose standardisation on widely used formats like CSV or SPSS, not realising that the **flexibility of these formats leaves the data insufficiently defined**. It does not help when **different sub-fields of a discipline are using the same terms to mean different things** (e.g. social sciences and humanities) or when there is **no standardisation of the way variables are coded**. Inter-disciplinary interoperability brings its own challenges: different repositories are using **different semantics for resolving persistent identifiers**, which makes it hard for machines to access the data. Some interdisciplinary standards like ORCID[9] are not **equally adopted in all disciplines**. In addition, solving findability and accessibility of data within a discipline by bringing the data together in a virtual research environment can result in a **larger silo** of data that no longer interoperates with other disciplines, emphasising the importance of community-specific solutions **[Recommendation 2]**.

Both findability and reusability require **metadata**. The widest reported technical problem with metadata is that there are **insufficient ways of automatically collecting and preserving** it. Currently, electronic lab notebooks are **either not flexible enough or are incapable to interface** with e.g. instrumentation that collects the data. While in one of the studies an overwhelming majority of

[7] Towards the Tipping Point for FAIR Implementation: https://doi.org/10.1162/dint_a_00049

[8] This section does not separately reference the documents from our reading list; it is a summary of all findings.

[9] https://orcid.org/

researchers report that they will only consider reusing a data set if it is very well documented, a similarly large percentage will be **put off by the prospect of having to document** their own data manually.

We encountered two related financial issues. First, it is very **hard to find dedicated funding for community-resources** over a longer period, covering e.g. changes in data standards. Second, many funders do **not allow researchers to budget long term service fees** that pay for data services beyond the lifetime of a project.

Social impediments

In different disciplines different reasons are brought up why the **FAIR principles do not apply** to data. This is often caused by **confusing FAIR with fully open** and freely accessible. In some cases the **high volume of data** (e.g. molecular sciences) is brought up. Elsewhere, the presence of **personal and sensitive data** (e.g. in the health sciences) has made many researchers think that FAIR, born out of the "open science" movement, did not apply to them. FAIR is also perceived to be unsuitable where **intellectual property protection** is essential due to the role of commercial parties (e.g. in engineering, health and plant sciences). Sometimes it is said that FAIR was made for quantitative data and not **qualitative data** (e.g. social sciences and humanities), or that it is not suitable for the study of **real world objects** because that is different from the study of digitised objects (e.g. humanities, but much less in natural history collections).

It is widely seen that researchers **do not see sufficient benefits** of FAIR data, and therefore are not willing to put in the efforts in implementing FAIR practices; this is sometimes phrased as **academic recognition coming primarily from papers** (explicitly mentioned in earth sciences) and not from data. In some cases, data is **not considered an autonomous research output**, but only supplementary to the paper. A related issue is that there is an academic **benefit of proposing and publishing new standards** over re-using existing ones.

We also see that researchers **do not think their data can be reused** for other research at all. In contrast, many feel that there would be **significant additional cost** incurred if data needs to become FAIR, because it is **hard to do** and a lot of **extra work is required**.

It is also observed that researchers are afraid that their data is **exploited by others:** they **fear being scooped** by others who run with the carefully collected data, or fear that the data will be misused by those who will **make commercial use** of it, who do **not understand** the data properly, or have **malicious intentions**.

In some fields, it is felt that it is **impossible to document data sufficiently** to allow other humans and machines to interpret it, and that human-human collaboration will therefore always be needed. We also observe that in different disciplines the **general resistance to change** in habitual processes is brought up.

Implementation of FAIR is sometimes impeded by **misunderstandings about copyright and licensing**. In life sciences researchers often think that data is **owned by the researcher**. In mathematics it is sometimes thought that putting something on a **website makes it public domain**.

Many of these arguments are caused by a widely observed **lack of sufficient knowledge and understanding of FAIR**: many researchers have never heard of the FAIR principles. It is also observed that **researchers do not have sufficient legal knowledge** to make data FAIR without proper legal support.

Many of these arguments against open or FAIR data are sufficiently addressed elsewhere; we will not repeat these here[10]. However, we want to make clear that FAIR is a journey that is taken step by step, and that the results of making data FAIR do not have to be perfect in order for them to be valuable **[Recommendation 1-2]**.

Technical solutions

When looking at the different disciplines it is important to recognise that some disciplines require different types of technical solutions to obtain the same benefits from FAIR data. For example "Findability" of data associated with a specific high-energy physics experiment may be sufficiently addressed if major search engines can find the instrument by name, whereas health researchers interested in a rare disease will need more Findability infrastructure to assemble information independently collected in many locations.

Generally, we observe that it becomes **easier to make data citable**; citing persistent identifiers becomes mainstream and many **repositories make it very easy to get a persistent identifier**, e.g. a DOI or Handle, for a data set.

Generic **repositories** in particular are augmenting their services to **support researchers in making data more FAIR**. The Core Trust Seal[11], for example, contains **guarantees for longevity of the data**, including fall-back contracts. Some long-existing, community-specific resources are slower to adapt their existing processes and standards.

In many fields there is **no shortage of data and metadata standards**; standards are becoming easily findable through resources like FAIRsharing[12]. Communities are getting together to **choose between different available standards**.

The role of semantics in interoperability is broadly recognised and **facilities for semantic interoperability are developed**, allowing better machine actionability of data. Good practices for semantic resources are being developed[13].

Some research disciplines are further along than others in implementing FAIR practice. In some cases this is due to a **long history of data sharing practice**, such as in astronomy and high-energy physics. Their large infrastructure, shared between researchers from many different institutes, have been designed with **data standardisation processes in mind**. In such disciplines, concrete, innate

[10] Concerns about opening up data, and responses which have proved effective: https://docs.google.com/document/d/1nDtHpnIDTY_G32EMJniXaOGBufjHCCk4VC9WGOf7jK4/edit#

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

[12] https://fairsharing.org/

[13] D2.2 FAIR Semantics: First recommendations; https://doi.org/10.5281/zenodo.3707985

demand for sharing and standardisation were decisive factors in their success stories. In these fields, the data is maintained by the infrastructures who have been collecting it.

There are practices that started as an effort in one discipline but could be readily generalised. For example, life sciences started **collecting and documenting the use of data and metadata standards** in BIOsharing; the realisation that this solved a problem of findability of standards that is also faced in other disciplines led to the development of FAIRsharing.

Life sciences have many **data-type specific repositories** which can offer more functionality for data re-users than generic repositories. This is a good model, but it may be hard to replicate for research fields where data types are less standardised. Also, each of these repositories requires sustained funding **[Recommendation 2]**.

Bringing together data and facilities for analysis into **Virtual Research Environments** increases findability and accessibility of the data (earth sciences). Related to this is the effort of **bringing the analysis to the data** instead of migrating the data to the place where they are to be analysed (e.g. earth sciences and life sciences); this approach solves problems with large data transfer as well as legal difficulties with off-premise copies.

Social enablers

Both publishers and research funders are in a position to **push for FAIR data sharing**. For funders this can be through mandates, as well as by **allowing projects to budget** for data management and data publishing. Funder's actions can be made effective by **monitoring** adherence **[Recommendation 7]**. Publishers can mandate data sharing, and can also **require authors to cite data** instead of just mentioning it.

A balance of penalties and rewards is needed for optimum impact. Rewards for data sharing that are mentioned in different places are **co-authorships** for the originators of data, or **being cited** as data authors. It is expected that the academic reward is in balance with the effort made in sharing the data (e.g. earth sciences). It is also suggested that data sharing should be incorporated into researcher's **performance evaluations [Recommendation 6]**.

The **disciplinary culture** is considered very important for data sharing: it is facilitated if **data sharing is the norm** in a discipline (e.g. astronomy). It can also help when a community is **organised around a virtual research environment** (e.g. earth sciences). Also, a **culture of collaboration** pushes data sharing along. Data from complex fields also push for data sharing because of the **pressure for verifiability**. A copyright policy that favours sharing data can also bring FAIR implementation forward.

Data sharing can be boosted by **increasing awareness** and through **education**. It helps if researchers know of **success stories**. Broad awareness also leads to peer visibility and **peer pressure**. Awareness can also be raised by the **availability of Research Data Management support** or through **Data Management Plan templates** that stress the importance of FAIR data. Researchers need to know that **FAIR data is not the same as open data** (many of these are mentioned in reports from social sciences and humanities).

Finally, it is easier to see the benefits of FAIR data when **collecting the data is either very expensive** or when there is only **a single chance of collecting an observation**.

It is interesting to note that the push for data sharing also results in a **push for better quality data in general**.

FAIR practices: a regional perspective

Regional FAIR Practices are strongly determined by national policies. In this section we give an overview of the commonalities and the regional differences observed. Overall, we found that within Europe, Western European countries, and in particular, the Netherlands, UK, France and Germany are in the lead when it comes to FAIR practice.

Main approaches in Europe

We observed eight main approaches towards introducing policies on FAIR practices in Europe, which we describe below with some representative examples. These approaches can be divided into three groups: national approach (National plan or policy (1a), expert or working groups developing policy usually on the request of the national government (1b)); funders' or infrastructure requirements (government-funded research (2a), national funder's policy (2b), national research data infrastructure requirements (2c)), and community approach (research institutions or research groups collaboration (3a), research integrity policies (3b), regional working groups enabling FAIR (3c)).

1a) National plan or policy: The Netherlands is one of the leaders in implementing FAIR principles in Western Europe. The Dutch *National Plan Open Science*[14] has an ambition for a consistent system to allow FAIR access to research data. The Plan is implemented through National Platform Open Science. The Netherlands is one of the few countries who have paid attention to monitoring and rewarding data sharing in their Plan. Similar national level framework approach is taken in Ireland through the *National Framework on the Transition to an Open Research Environment*[15] and in Norway through the Ministry of Education and Research's *National Strategy on Access to and Sharing of Research Data*[16]. These national policies often do not mention FAIR, but the approach contains all elements of FAIR. For example, both Serbia[17] and Slovenia[18] describe in their strategies how and

[14] van Wezenbeek, W.J.S.M., Touwen, H.J.J., Versteeg, A.M.C., and van Wesenbeeck, A. (2017). National Plan Open Science. Ministerie van Onderwijs, Cultuur en Wetenschap, 2017. https://doi.org/10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65.

[15] National Open Research Forum (July, 2019). National Framework on the Transition to an Open Research Environment. http://norf-ireland.net/wp-content/uploads/2019/07/NORF_Framework_10_July_2019-2.pdf

[16] Norwegian Ministry of Education and Research (2018). National Strategy on Access to and Sharing of Research Data. https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/en-gb/pdfs/national-strategy-on-access_summary.pdf

[17] Ministry of Education, Science and Technological Development of the Republic of Serbia. (July, 2018). Open Science Platform. http://open.ac.rs/svevesti/87328781babfe70aad60429fad8f4feb/Open-Science-Policy-Serbia.pdf

[18] Government of the Republic of Slovenia. (3 September, 2015). National Strategy of Open Access to Scientific publications and research data in Slovenia 2015–2020. https://www.gov.si/assets/ministrstva/MIZS/Dokumenti/ZNANOST/Strategije/National-strategy-of-open-access-to-scientific-publications-and-research-data-in-Slovenia-2015-2020.pdf

when research data should be made available, as well as which repositories and licences should be used.

1b) Policy recommendations of national level workgroups: such workgroups give recommendations and advice on principles for development of national open science policies. In Austria, the Open Science Network Austria (OANA) WG has developed recommendations for a national open science strategy[19], which includes FAIR recommendations. A similar approach has been taken also in Baltic and Eastern European countries (Estonia, Latvia, Slovakia).

2a) Compliance requirements for government-funded research: France has set a goal in its National Plan for Open Science[20] to ensure that data produced by government-funded research become gradually compliant with the FAIR Data Principles and that they are preserved and, whenever possible, open to all. The same principle is applied in Norway - the Research Council of Norway Policy[21] for open access to research data has been based on the FAIR Principles.

2b) National funders' policy/requirements: Belgian federal funder's BELSPO Open Research Data policy complies with FAIR principles.[22]

2c) Requirements of national research data infrastructures: National Research Data Infrastructures [23] (NFDI) funded by Germany's federal and state governments require that all data preserved in NFDI is managed in accordance with FAIR principles.[24] A similar approach is taken also in Italy where the Italian Computing and Data Infrastructure (ICDI[25]) is leading in FAIR practices in order to establish a nationally coordinated strategy towards FAIR.

3a) Multi-stakeholder approach to requirements: In the UK, the Concordat on Open Research Data [26] has been developed by a multi-stakeholder group and has been signed by the higher education funding council, one private funder (Wellcome Trust), several national research funders, and the umbrella group of UK universities. The Concordat is not considered a government document but rather a community output. The Concordat does not focus specifically on FAIR but the content is

[19] Open Science Network Austria OANA. (2020). Recommendations for a National Open Science Strategy in Austria of the Open Science Network Austria OANA written by the working group "Open Science Strategy".
https://oana.at/fileadmin/user_upload/k_oana/dokumente/Entwurfv1.1-EmpfehlungenOS-OANA.pdf

[20] National plan for Open Science. (4 July, 2018).
https://www.ouvrirlascience.fr/national-plan-for-open-science-4th-july-2018/

[21] The Research Council of Norway (March, 2020). The Research Council Policy for Open Science.
https://www.forskningsradet.no/siteassets/tall-og-statistikk-seksjonen/apen-forskning/nfr-policy-open-science-eng.pdf.

[22] The Federal Science Policy Office (BELSPO). (3 December, 2019). Open Research Data mandate.
https://www.belspo.be/belspo/openscience/doc/ORD_Policy_Dec2019.pdf

[23] German Research Foundation. National Research Data Infrastructures (website)
https://www.dfg.de/en/research_funding/programmes/nfdi/index.html

[24] German Research Foundation. (May, 2020). Guidelines for Consortia National Research Data Infrastructure (NFDI)https://www.dfg.de/formulare/nfdi100/nfdi100_en.pdf

[25] Proudman, V., Sveinsdottir, T., & Davidson, J. (2020). *An Analysis of Open Science Policies in Europe v5*. Zenodo. https://doi.org/10.5281/zenodo.3689450

[26] UK Research and Innovation. (28 July, 2016). Concordat on Open Research Data.
https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf

aligned with the FAIR principles. In addition, the Open Research Data Task Force[27], which builds its recommendations on the principles set out in the Concordat, argues for adherence to FAIR principles for sharing data in the UK. A multi-stakeholder approach is taken in Finland's Declaration for Open Science and Research (Finland) 2020-2025.[28] All signed organisations and research communities in Finland accept that the management of research data is based on FAIR principles.

3b) Research integrity policies: FAIR data principles are often referred to in national or institutional codes of conduct for research integrity. The Netherlands Code of Conduct for Research Integrity[29] asks researchers to contribute to FAIR data and tasks research institutions with ensuring that research data is open and accessible in accordance with the FAIR principles. Similar approaches are also used in countries or institutions without any official national policy on FAIR or Open Science, for example in Estonia[30].

3c) Regional approach - Nordic and Baltic countries have taken a collaborative approach to FAIRification of repositories through the EOSC Nordic project[31]. Goals of this project are to identify the region's research data repositories, evaluate and improve their FAIRness and to landscape Open Science policies in Nordic and Baltic countries.

Outside Europe

Countries outside Europe are also actively involved in developing their open science and FAIR policies. The Australian Research Data Commons (ARDC[32]) supports and encourages initiatives that enable making data and other related research outputs FAIR, including policy development.[33] With the policy statement, Australia confirms that all their publicly funded research outputs will be FAIR.[34] New Zealand's eResearch2020 is a nationally coordinated multi-stakeholder programme for developing a national strategic approach to research data in New Zealand.[35] The United States in a 2013 Directive from the White House Office of Science and Technology Policy (OSTP) required

---

[27] Open Research Data Task Force. (July, 2018). Realising the potential. (Final report). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf

[28] Open Science Coordination in Finland, Federation of Finnish Learned Societies (2020). Declaration for Open Science and Research 2020–2025, 2nd edition. DOI https://doi.org/10.23847/isbn.9789525995213

[29] KNAW; NFU; NWO; TO2-federatie; Vereniging Hogescholen; VSNU. (2018). Neatherland's Code of Conduct for Research Integrity. DANS. https://doi.org/10.17026/dans-2cj-nvwu

[30] Centre for Ethics, University of Tartu (2017). Estonian Code of Conduct for Research Integrity. https://www.eetika.ee/sites/default/files/www_ut/hea_teadustava_eng_trukis.pdf

[31] EOSC-Nordic. (2020). European Commission Horizon 2020 project no. 857652. (website). https://www.eosc-nordic.eu/

[32] Australian Research Data Commons. (website). https://ardc.edu.au/

[33] Australian Research Data Commons (2020). FAIR principles. (website). https://ardc.edu.au/collaborations/fair-principles/

[34] Australian FAIR access working group (2020). Policy Statement on F.A.I.R. Access to Australia's Research Outputs. https://www.fair-access.net.au/fair-statement

[35] NeSI, REANNZ and NZGL. (March, 2016). eResearch 2020. National Research Data Programme. http://www.eresearch2020.org.nz/wp-content/uploads/2016/03/eResearch2020_NationalResearchDataPrograme_S.pdf

Public Access to Federally Funded Research Outputs[36]. This work is an ongoing federal effort to support the advancement of open science and to make federally funded research outputs available.[37]

General commonalities, differences, and gaps in Europe

Commonalities

- It is usually funders, institutions and research groups that introduce FAIR policies rather than governments. In many countries, national funding agencies are the main actors implementing open science strategies (this is noticeable all over Europe, with no regional differences).
- FAIR is mainly part of communities' practice rather than of national policies. Various research groups and disciplines are doing FAIR and they also mention FAIR in their proceedings.
- FAIR is sometimes confused with open data.
- Most Open Science national policies or recommendations require managing and sharing research data. There is no difference in this between various regions of Europe.
- Most European countries have established National Infrastructure Roadmaps which often contain research data infrastructure and recommendations on data management, preservation and usability.
- Various studies across Europe have shown that only few countries have an official Open Science policy that refers to FAIR data. Most Open Science and Access national policies have implications to FAIR principles without mentioning these explicitly. Policies cover preservation, accessibility, reusability, machine readability and other principles of FAIR. In Europe, at the beginning of 2020, only six countries had national Open Science policies where FAIR is mentioned. These included the policies from the Netherlands, France, the UK, Finland, Spain and Ireland. At the same time, more than 15 countries in Europe had Open Science national policies in place.[38]

Differences

- Main FAIR implementers are Western European countries. More focus needs to be placed on Eastern Europe, specifically in the area of FAIR[39], given that less national Open Science and FAIR policies have been adopted in the Eastern European and Baltic countries **[Recommendation 7**].

[36] Executive Office of the President. Office of Science and Technology Policy. (22 February, 2013). Increasing Access to the Results of Federally Funded Scientific Research. (Memorandum). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[37] Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting From Federally Funded Research. (2020, February 19). Federal Register. https://www.federalregister.gov/documents/2020/02/19/2020-03189/request-for-information-public-access-to-peer-reviewed-scholarly-publications-data-and-code

[38] Proudman, V., Sveinsdottir, T., & Davidson, J. (2020)., ibid.

[39] van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A., Kirkpatrick, C. and Mons, B., 2020. Towards the Tipping Point for FAIR Implementation. *Data Intelligence*, 2(1-2), pp.264-275. https://www.mitpressjournals.org/doi/full/10.1162/dint_a_00049

- *FAIRsFAIR FAIR D3.1 FAIR Policy Landscape Analysis*[40] survey showed that Western European countries are more active in Open Science and are more advanced in implementing FAIR guidelines (majority of survey respondents are from Western European countries). However, a more detailed look into Eastern European Open Science recommendations (there are few policies) reveals that these countries also recommend research data to be open, available and reusable. These differences between Western and Eastern European countries might be explained with the EU funding of various FAIR-related projects which are often led by Western European countries (GOFAIR, FAIRsFAIR), and in which Eastern European countries do not always participate.

Gaps

- Focusing only on the term FAIR is limiting the understanding of activities that are taking place in Europe that are advancing FAIR. Often FAIR is not mentioned, but activities are enabling implementation of FAIR principles in practice. It can be observed that countries are moving towards FAIR; however, the term FAIR is not widely spread yet.
- Studying FAIR and Open Science policies is not enough to landscape the work that is done. The mapping should be wider by including research integrity activities, teaching and training, and also by including actions taking place on institutional and discipline level **[Recommendation 7]**.

[40] Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., & Whyte, A. (2019). D3.1 FAIR Policy Landscape Analysis. https://doi.org/10.5281/zenodo.3558173

FAIR practices for other research objects

In the original FAIR principles paper[41] the authors state:

> *"...it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability."*

The majority of reports and studies on FAIR practice focus on research data; when other digital research objects are mentioned, it is in the role of supporting FAIR data, e.g. tools to enable FAIRification[42], such as the use of Data Management Plans (DMPs) and software to improve the data processing steps required before publication. This section of the report looks at the published practice and work to define better guidance to make other research objects FAIR in their own right.

FAIR Digital Research Objects

For the purposes of this report, we consider FAIR Digital Research Objects to include anything which is a direct component of the research process, e.g. software, workflows, executable notebooks, DMPs. This might also include research objects which are originally either physical (e.g. samples) or conceptual (e.g. protocols[43,44]) but have a directly referenceable digital form; but we do not include these in the current version of the report due to a lack of published practice. Many research objects are discipline-specific, which means that FAIR guidance and practice will also be discipline-specific.

We are excluding indirect components of the research process, such as teaching and training materials, from this report but note that significant progress has been made under the banner of "Open Educational Resources" (OER) that is complementary to the adoption of FAIR. The EOSCpilot project considered how standards developed for OER may be applied towards making training

[41] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1). https://doi.org/10.1038/sdata.2016.18

[42] Thompson, M., Burger, K., Kaliyaperumal, R., Roos, M., & da Silva Santos, L. O. B. (2020). Making FAIR Easy with FAIR Tools: From Creolization to Convergence. Data Intelligence, 2(1–2), 87–95. https://doi.org/10.1162/dint_a_00031

[43] https://www.protocols.io/

[44] https://protocols.scienceexchange.com/

materials more FAIR.[45] Initiatives to catalogue these materials as FAIR resources are being led by e.g. ELIXIR,[46] ENVRI-FAIR,[47] and the EOSC Board WG Training and Skills.[48]

Current Practice

It is clear that adoption of FAIR practice for other research objects lags behind research data, yet evidence from the number of software deposits in repositories[49] and registries[50] with associated metadata and identifiers suggests that many research objects should be more findable and accessible. There is some evidence[51] of how different ESFRIs are making other types of outputs more findable and accessible as part of a broader aim of making their catalogues and repositories FAIRer, including EPOS aggregating information of about 400 elements (data, data product, software and services) to improve findability and CLARIN developing distributed workflow frameworks with harmonised metadata descriptions to improve interoperability and reusability.

The Turning FAIR into Reality report advocates that DMPs should be FAIR outputs in their own right. Making DMPs 'machine-actionable' means making their content findable and accessible, exchanging that content with other systems in standardised, interoperable ways, and potentially reusing that content. A standard for exchanging DMP content[52] has demonstrated the effective exchange of DMP data across several connected platforms[53].

However, most of the published practice, guidance and policy on other research objects concerns software, workflows and computational (executable) notebooks.

Software

Historically, there has been a wide spectrum of practice in publishing and sharing research software (including applications, scripts, tools, libraries, APIs and services). A previous lack of formalisation and standards means that even within disciplines, practices may vary considerably.

---

[45] Whyte, A., Leenarts, E., de Vries, J. et al. (2019) Strategy for Sustainable Development of Skills and Capabilities, EOSCpilot D7.5
https://eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities

[46] Garcia L, Batut B, Burke ML, Kuzak M, Psomopoulos F, Arcila R, et al. (2020) Ten simple rules for making training materials FAIR. PLoS Comput Biol 16(5): e1007854.
https://doi.org/10.1371/journal.pcbi.1007854

[47] https://trainingcatalogue.envri.eu/

[48] Kuchman, I. 'Building competence and capabilities for EOSC' (blog article, 30/032020)
https://www.eoscsecretariat.eu/news-opinion/competence-capabilities-eosc-skills-training

[49] Fenner, M. (2019). Jupyter Notebook FREYA PID Graph Key Performance Indicators (KPIs) (Version 1.1.0). DataCite. https://doi.org/10.14454/3BPW-W381

[50] E.g. Astrophysics Source Code Library https://ascl.net/ and DOE Code https://www.osti.gov/doecode/

[51] Wittenburg, P., de Jong, F., van Uytvanck, D., Cocco, M., Jeffery, K., Lautenschlager, M., Thiemann, H., Hellström, M., Asmi, A., & Holub, P. (2020). State of FAIRness in ESFRI Projects. Data Intelligence, 2(1–2), 230–237. https://doi.org/10.1162/dint_a_00045

[52] Walk, P., Miksa, T., & Neish, P. (2019). RDA DMP Common Standard for Machine-actionable Data Management Plans. Research Data Alliance. https://doi.org/10.15497/RDA00039

[53] https://rda-dmp-common.github.io/hackathon-2020/

The open source software community aims to allow anyone to inspect, modify and enhance software. They have developed practices and recommendations that align with FAIR principles, and which are increasingly used by researchers as open source licensing of research software becomes more common. For example, by following simple recommendations for making research software open[54] (make code public, add to registries, use open source license) it is possible to make software more findable, accessible and reusable. The practice of depositing software in an archive (for instance, when publishing a paper) is increasing due to changes in journal policies[55]. However, despite availability of guidance on publishing software[56], this is still not commonplace. In Zenodo, for instance, only 3.24% of all software DOIs registered are traceably cited at least once, and most are self-citations[57]. A study on GitHub repositories referenced in publications show clear differences in the reusability of the software[58] with 23.6% not having a license and readme - two basic indicators of reusability.

Most of the published work[59,60,61,62] on FAIR suggests that whilst the FAIR foundational principles can apply to software, the guiding principles require translation for software; though how much is still unclear. The paper "Towards FAIR principles for research software"[63] reviews previous work on applying the FAIR principles to software and suggests ways of adapting the principles to a software context. They argue that software is different from data: it is a tool to do something (executable); it is built by using other software (implements multi-step process, coordinates multiple tasks), it has complex dependencies and has a short life cycle with frequent need of versioning (including

---

[54] Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., … Crouch, S. (2017). Four simple recommendations to encourage best practices in research software. F1000Research, 6, 876. https://doi.org/10.12688/f1000research.11407.1

[55] E.g. BMC policy: https://www.biomedcentral.com/getpublished/writing-resources/structuring-your-data-materials-and-software

[56] Jackson, M. (2018). Software Deposit: Guidance For Researchers. Zenodo. https://doi.org/10.5281/ZENODO.1327310

[57] van de Sandt, S., Nielsen, L., Ioannidis, A., Muench, A., Henneken, E., Accomazzi, A., Bigarella, C., Lopez, J. and Dallmeier-Tiessen, S., 2019. Practice Meets Principle: Tracking Software And Data Citations To Zenodo Dois. [online] arXiv.org. Available at: https://arxiv.org/abs/1911.00295 [Accessed 18 June 2020].

[58] Whitaker, K., O'Reilly, M., , Isla, & Hong, N. C. (2018). Softwaresaved/Code-Cite: Sn-Hackday Version. Zenodo. https://doi.org/10.5281/ZENODO.1209095

[59] Chue Hong, N., & Katz, D. S. (2018). FAIR enough? Can we (already) benefit from applying the FAIR data principles to software? https://doi.org/10.6084/M9.FIGSHARE.7449239.V2

[60] Erdmann, C., Simons, N., Otsuji, R., Labou, S., Johnson, R., Castelao, G., Boas, B. V., Lamprecht, A.-L., Ortiz, C. M., Garcia, L., Kuzak, M., Martinez, P. A., Stokes, L., Honeyman, T., Wise, S., Quan, J., Peterson, S., Neeser, A., Karvovskaya, L., … Dennis, T. (2019). Top 10 FAIR Data & Software Things. Zenodo. https://doi.org/10.5281/ZENODO.2555498

[61] Aerts, P. J. C. (2017). Sustainable Software Sustainability - Workshop report. Data Archiving and Networked Services (DANS). https://doi.org/10.17026/DANS-XFE-RN2W

[62] Doorn, P. (2017). Does it make sense to apply the FAIR Data Principles to Software? https://indico.cern.ch/event/588219/contributions/2384979/attachments/1426152/2189855/FAIR_Software_Principles_CERN_March_2017.pdf

[63] Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. L., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2020). Towards FAIR principles for research software. Data Science, 3(1), 37–59. https://doi.org/10.3233/DS-190026

dependencies). Some of these characteristics also apply to data. However, the variety of software and its publishing and distribution channels, and the necessity to document dependencies and describe data formats, poses a challenge when adapting the current FAIR principles.

Recent recommendations for FAIR software[64] note that "at present research software is typically not published and archived using the same practices as FAIR data, with a common vocabulary to describe the artefacts with metadata and in a citable way with a persistent identifier". The majority of software is effectively "self-published", through project websites or code repositories such as GitHub and Bitbucket, rather than going through a deposit and curation step, as is the case with publishing data in a digital repository. The use of discipline-specific, community-maintained catalogues and registries (e.g. in astronomy[65], biosciences[66], geosciences[67]) can make software more findable and accessible if software is registered in them. Increasing incentives for publishing software with good metadata, such as improved acceptance of software citation[68] and the ability to make software more discoverable through search engines through improved annotation will help to increase the findability and accessibility of software. However, this does not address the issue of information loss in the scholarly publishing system[69], where the metadata provided by software authors is removed or incorrectly passed on, which hinders the ability of infrastructure and tools used by research to help make software FAIR by degrading citations, credit and discoverability.

Services

Software is often used to provide web services to process or analyse data. These services are typically domain-specific and some communities have identified the need for FAIR services. In the marine sciences, properly structured metadata to aid findability, along with provision of services via uniform and compatible encodings using community-adopted standards to aid accessibility, will be required to support machine-based processing of data flows[70]. In biodiversity, a digital object architecture has been proposed as an approach, building on the use of community-specific metadata registries[71]. GO-FAIR suggests using the 'hourglass model' to support 'The Internet of FAIR Data and Services'[72], where (similar to the architecture of the internet which has network protocols, e.g. IP, at the "neck" in the middle of the hourglass as an abstraction / spanning layer between the proliferation of applications above and physical networks below) a small set of core pieces -

[64] Hasselbring, W., Carr, L., Hettrick, S., Packer, H., & Tiropanis, T. (2020). From FAIR research data toward FAIR and open research software. It - Information Technology, 62(1), 39–47. https://doi.org/10.1515/itit-2019-0040

[65] ASCL: https://ascl.net/

[66] BioTools: https://bio.tools/

[67] OntoSoft: https://www.ontosoft.org/

[68] Smith, A. M., Katz, D. S., & Niemeyer, K. E. (2016). Software citation principles. PeerJ Computer Science, 2, e86. https://doi.org/10.7717/peerj-cs.86

[69] Nielsen, L. H., & Van De Sandt, S. (2019). Tracking citations to research software via PIDs. ETH Zurich. https://doi.org/10.3929/ETHZ-B-000365763

[70] Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., de Bruin, T., … Zhao, Z. (2019). Ocean FAIR Data Services. Frontiers in Marine Science, 6. https://doi.org/10.3389/fmars.2019.00440

[71] Lannom, L., Koureas, D., & Hardisty, A. R. (2020). FAIR Data and Services in Biodiversity Science and Geoscience. Data Intelligence, 2(1–2), 122–130. https://doi.org/10.1162/dint_a_00034

[72] https://www.go-fair.org/resources/internet-fair-data-services/

persistent identifiers and mapping tables - are agreed to support FAIR data, tools and services. In all cases, these approaches are still on the path to adoption and maturity.

The FAIRsFAIR Assessment report on 'FAIRness of services'[73] identified that "mapping of the 15 FAIR principles [...] to data services would [...] probably not deliver actionable insights of real and lasting value" and that "there is limited tangible guidance on how to 'make services FAIR'". The ongoing work in FAIRsFAIR will be developing a Data Services Assessment Framework that will include actionable recommendations that service providers need to make incremental improvements to their services to support the emergence of a FAIR ecosystem.

Workflows

The history of sharing workflows dates back to before the publishing of the FAIR principles. Initiatives such as the Galaxy Toolshed[74] and myExperiment[75] in the life sciences and ArcGIS Catalog [76] in geosciences have made computational and data processing workflows more findable, accessible and reusable, before the FAIR principles were conceived.

Most current publications on FAIR workflows suggest policies and processes to improve the FAIRness of workflows. These include the use of persistent identifiers (PIDs) and machine learning to improve classification[77]; and better conventions for naming workflows alongside registration in specialised repositories[78]. A common theme is that the same challenges faced when attempting to apply the FAIR guiding principles to software apply to workflows and executable notebooks; their characteristics mean that they are similar to software artefacts. Another challenge for workflows is that automated annotation and description strategies and tools are required because the burden of creating and maintaining metadata for workflows is much higher than for data.

Executable notebooks

A significant cultural change has occurred in the last five years, with more research[79] being disseminated through executable notebooks (most commonly Jupyter Notebooks). In the geosciences, domain-specific software repositories and better specification of software location, license and citation are suggested as ways of making research software findable and accessible,

[73] Koers, H., Gruenpeter, M., Herterich, P., Hooft, R., Jones, S., Parland-von Essen, J., & Staiger, C. (2020). Assessment report on 'FAIRness of services'. https://doi.org/10.5281/ZENODO.3688762

[74] https://galaxyproject.org/toolshed/workflow-sharing/

[75] https://www.myexperiment.org/

[76] https://pro.arcgis.com/en/pro-app/help/analysis/geoprocessing/share-analysis/create-a-geoprocessing-package.htm

[77] Weigel, T., Schwardmann, U., Klump, J., Bendoukha, S., & Quick, R. (2020). Making Data and Workflows Findable for Machines. Data Intelligence, 2(1–2), 40–46. https://doi.org/10.1162/dint_a_00026

[78] Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., … Schober, D. (2020). FAIR Computational Workflows. Data Intelligence, 2(1–2), 108–121. https://doi.org/10.1162/dint_a_00033

[79] E.g. the LIGO Project: https://losc.ligo.org/tutorials/

along with using containers to make software easier to reuse, to create “Geoscience papers of the future” combining data, code and narrative[80].

Considerable progress has been made on tooling and services to help make executable notebooks findable, accessible and reusable, by providing DOIs to identify them, reproducible environments to run them (Binder[81], CodeOcean[82]) or to export them to other publishing formats. This has been supported by documentation and training that has aided adoption. One study has analysed the FAIRness of Jupyter notebooks in the Astrophysics Data System, with 37 of 91 papers publishing openly accessible Jupyter notebooks containing detailed research procedures, associated code, analytical methods, and results. However, practices for mentioning, storing, and providing access to the notebooks varied greatly across papers[83].

Commonalities and Gaps

Analysis of existing practice and guidance reveals a number of commonalities shared across software, workflows and executable notebooks in relation to improving adoption of the FAIR principles - these should continue to be addressed:

- **Identifiers** are seen as a key requirement to making research objects findable and accessible. However, uptake of suitable persistent identifiers with associated metadata, though increasing, is still relatively low. **[Recommendation 2, 5]**
- **Specialist repositories and catalogues** are often suggested to improve the FAIRness of software and workflows. These improve the quality of the metadata associated with other research objects, but require additional effort from developers and curators. The adoption of these infrastructures is often related to their use for other research objects in particular domains. **[Recommendation 1, 2, 4, 5]**
- **Publishing of software** is different from publishing of data. Because the community norms of distributing software do not currently include the use of FAIR repositories, there is less cohesion around metadata. Making the metadata curation part of the process of assigning identifiers may help. Changes in code repository infrastructure, such as support for keywords/topics[84], will make it easier to automatically harvest and collate such information, which will make it easier to implement “metasearch” engines to improve the findability of software, workflow and services without them needing to be deposited in repositories. **[Recommendation 2, 3, 5, 6]**

---

[80] Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R., & Yu, X. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. Earth and Space Science, 3(10), 388–415. https://doi.org/10.1002/2015ea000136

[81] https://mybinder.org/

[82] https://codeocean.com/

[83] Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). Presented at the 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). https://doi.org/10.1109/jcdl.2017.7991618

[84] https://help.github.com/en/github/administering-a-repository/classifying-your-repository-with-topics

- **Authorship** - including citation and credit policies - are often mentioned as a method of providing incentives to improve FAIRness. Publishers, journals and conferences have shown a willingness to provide better support for this. **[Recommendation 1, 3, 5, 6]**

There are also some key gaps, where work is only just beginning:

- **Executable papers** combine elements of data, software, workflow and paper. It is still unclear how practice around making executable papers FAIR might proceed, though there is a proposed RDA effort to examine this. **[Recommendation 3, 5]**
- **Metrics for FAIR software,** as currently proposed, combine metrics based on FAIR data metrics with metrics based on software quality metrics. This will need to be clarified, in particular to identify which metrics will best help adoption of FAIR for software. **[Recommendation 5, 7]**
- **Studies on the adoption of FAIR** for other research objects are rare. Most published work looks at limited case studies, or proposes recommendations on how to apply FAIR principles, rather than measuring the success of these recommendations. **[Recommendation 7]**

Applying FAIR principles to the context of specific communities requires adoption/translation. This need is more obvious in the case of other digital research objects such as software. The relative importance of the FAIR foundational principles will depend on the goals, priorities and open science / open research culture of the community. Funder and publisher mandates will also have a key role in improving FAIR practice, as most of what has been identified in this section has resulted in requirements to share code as a prerequisite for publication.

In 2020, a joint RDA/FORCE11/ReSA working group has been setup on FAIR for Research Software (FAIR4RS)[85], which will be reviewing and, if necessary, redefining FAIR guiding principles for software and related computational code-based research objects. We expect this to be the community forum for taking forward the FAIR principles for software, services and workflows.

[85] https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg

Addressing differences in FAIR maturity between communities

The adoption of FAIR principles is increasing, but for policymakers and research funders it is important to acknowledge that in reality researchers know little about FAIR. In 2018, 60% of surveyed researchers had never heard of FAIR and only a fraction understood what the FAIR principles meant. Even awareness of an usually powerful tool, funders' expectations, was only 30%.[86] Reality is even more harsh: among the respondents of such surveys there is usually a bias towards disciplines, countries and groups that already have better awareness. For example, bio- and natural sciences are significantly overrepresented, constituting almost the whole sample of disciplines where FAIR Guiding Principles have been properly implemented. This has also other implications: recommendations, standards and, more importantly, expectations regarding FAIR are based largely on experiences and expertise from these most successful and engaged communities.[87] This section will discuss how differences in maturity between different research communities affect FAIR practices and why they should be taken into account when policymakers and research funders make decisions on research data, on the possible adoption of FAIR practices, and on the allocation of funding to support these activities.

It is very important to understand the reasons for the large differences between research communities and groups that are advanced in practising FAIR data and those that are not, as this has strong implications for the EOSC. Why do some communities already now practise FAIR data and why do others not? Why simply demanding that communities work harder to implement FAIR practices won't suffice? We need to understand this in order to enable all researchers from all communities to participate in the EOSC, and to implement FAIR data and benefit from it.

Importance of internal drivers

Our observations support a conclusion that the successful implementation of FAIR practices in a particular community is usually a result of bottom-up initiatives. These initiatives typically arise from concrete demands for each other's data. Development of awareness inside the group on what data is important and should be shared is a crucial step here. In communities where there is internally a high level of reuse, researchers are intrinsically motivated to share their data. This motivation is crucial, but it is not sufficient for a community to establish FAIR data routines. Effective data sharing requires certain standards for findability, availability and interoperability and some thought on reusability. There are thus many factors involved, such as the type of data in question (how easy is it to standardise? are there any legal challenges such as GDPR or IPR issues?), the degree of organisation and international cooperation a community represents (are there any community governance structures? is there a forum where such things can be discussed and decided?), the financial resources of the community (who will pay for the infrastructure? who will lead and sustain these efforts?), etc. The availability of standards, methodologies and infrastructure for FAIR data in a

[86] State of Open Data 2018. TU Delft survey.
[87] https://www.mitpressjournals.org/doi/full/10.1162/dint_a_00049

given community will depend on all of these, as well as on the hard work of community members to develop these.

As a result, the divides between more and less advanced groups do not strictly follow discipline boundaries, but they also exist inside disciplines and subdisciplines, depending on types of data collected, country or region where a researcher is based, or even age groups among researchers. In the end these different factors boil down to whether there is genuine demand for FAIR practices and whether the effort required to achieve them is reasonable; if yes, then a culture that embraces those practices may develop. But if a community lacks strong internal motivation, then often the barriers to FAIR data are considered too high and no cultural change will occur.

Top-down approaches need to take into account community needs

As mentioned above, expectations regarding FAIR practices are based on the experiences of communities which have successfully embraced the FAIR principles. Funders and policymakers tend to take these experiences and then transform them into expectations and solutions applicable to all research communities. At this point, the bottom-up success stories are transformed into top-down endeavours which have drastically lower success rates. For communities that are not familiar with FAIR, such demands increase the feeling of alienation, in particular if these communities/groups haven't yet found their innate demand for FAIR data. Tools that are recommended for use are often developed to different needs in different communities and may feel unfamiliar in different fields and their use inconsequential. In addition, if there is no internal awareness on what data should be FAIR, demands might be interpreted so that everything should be made FAIR, which is both a daunting and likely impossible (if not undesirable) task to begin with. When faced with demands from research funders and policymakers, some might adopt these principles and demands, but only superficially, and, without proper support, they might end up publishing data that is not FAIR.[88]

The change of direction from bottom-up to top-down also has other implications. When systems have been built from the bottom-up, they have evolved naturally, building up infrastructure, services and expertise that are truly necessary for the successful implementation of FAIR practices in the particular community. When applied from top-down, these communities might only have the most general level of support from general data experts, if any support is available at all, and not from field-specific data stewards that are essential for successful implementation of FAIR and take-up by researchers [**Recommendation 1**]. If the group as a whole doesn't yet have a shared understanding for issues related to FAIR, it direly needs specialists who have such an understanding and can support the rest of its members on their specific field of research.

It is a natural conclusion that we cannot simply wait for all the disciplines and groups to find FAIR on their own. While translating bottom-up experiences and success stories into top-down policies and expectations is necessary, it is essential that such policies and expectations truly reflect community needs and practices in order to be successfully implemented. They also need to take into account what made bottom-up success possible and why the top-down approach faces

[88] https://datascience.codata.org/articles/10.5334/dsj-2017-016/

difficulties. Here policymakers and especially research funders and institutions are seen as crucial contributors to change as their demands have the power to influence how researchers behave.[89] But in driving these changes it is recommended that funders and research institutions take into consideration how much work has to be done on grassroot level to engage those researchers who are not yet familiar with FAIR **[Recommendation 1-3]**. To succeed, the audience as a whole has to be understood and serviced **[Recommendation 1-3]**. They have to be incentivised **[Recommendation 4-6]**. It means investing also in research support services and raising awareness of their pivotal role, not only giving demands and recommendations **[Recommendation 1, 7]**. It is also crucial to always bear in mind that FAIR is not binary: FAIR/unFAIR, but a wide spectrum. If we expect to project practices from well developed communities and success stories suddenly on everybody, many will be overwhelmed as they have not had time to follow the long path towards it. It is a path where every step is valuable and every step needs support, services and training **[Recommendation 1-4]**.

[89] State of Open Data Report, 2019. https://doi.org/10.6084/m9.figshare.9980783.v1

Recommendations for improving FAIR practices

In order to ensure widespread benefits of the EOSC, improvements in FAIR practices are necessary. The first essential step to achieve this is for the communities to develop a shared understanding of their internal needs for FAIR practices. Shared understanding could in turn motivate the development of agreed methodologies, standards, tools, policies and infrastructures. FAIR data is a goal that cannot be achieved in one leap. Rather, it is a journey and each step, even a small one, is essential and valuable.

Therefore, to facilitate widespread adoption of FAIR practices, all these steps need to be incentivised and we make the following seven key recommendations:

1. Fund community-specific activities to raise awareness about the FAIR principles, their prerequisites and their benefits, and to establish dedicated support for their practical implementation.
2. Provide financial support for communities to develop, adopt, implement and maintain community standards, methodologies, tools and infrastructure.
3. Provide funding for training and for sustaining coordination fora and expert groups to help communities with standard development.
4. Incentivise development of community governance.
5. Recognise that FAIR guidelines will require translation for other digital objects and support such efforts.
6. Reward and recognise improvement of FAIR practice.
7. Develop and monitor adequate policies for FAIR data.

These recommendations are explained below, indicating the key stakeholder groups tasked with applying these recommendations, and providing a short rationale and practical examples.

Recommendation 1: Fund community-specific activities to raise awareness about the FAIR principles, their prerequisites and their benefits, and to establish dedicated support for their practical implementation.

Stakeholders: EOSC, Research funders, Institutions
Rationale: Community-specific actions are needed because arguments and solutions which work for one community might not be the key drivers for another. Awareness raising and providing dedicated community-specific support takes time and effort and thus such actions need to be financially supported. Funding pilot projects might be a useful mechanism to facilitate this.
Example: An initial pilot at TU Delft to fund data stewards with disciplinary knowledge helped communities realise the importance of FAIR practices, foster best practices and

prompted them to appoint their data stewards as permanent members of staff[90]. Funding similar pilots could help other communities see the value of FAIR practices and drive the internal need for improvement.

Recommendation 2: Provide financial support for communities to develop, adopt, implement and maintain community standards, methodologies, tools and infrastructure.

Stakeholders: EOSC, Research funders
Rationale: It is difficult for communities to work without funds, on a best effort basis. The development of standards, methodologies and tools takes commitment and time[91] . However, this phase is essential for putting FAIR principles into practice. While it is important that community members actively contribute to standard development, leading such work requires dedicated resources. Funding of adoption efforts is also crucial, in order to avoid unnecessary overproliferation of standards and to facilitate alignment and interoperability between various communities. Implementation of standards also requires appropriate tools and infrastructure (e.g. databases, repositories), tailored to community needs, and the development of these also needs to be funded. Standards, tools and infrastructure also have to be sustainably maintained and regularly revised to avoid depreciation, and this can only happen if communities see the value of such standardisation, are incentivised to do such work, and receive the necessary funding for this.
Example: The Joint Programme on Wind Energy of the European Energy Research Alliance (EERA JPWind) received funding from the European Commission which allowed it to lead concentrated efforts which culminated in successful development of taxonomy and metadata for the wind energy sector[92].

[90] Cite: Plomp, Esther, Nicolas Dintzner, Marta Teperek, and Alastair Dunning. 2019. "Cultural Obstacles to Research Data Management and Sharing at TU Delft". Insights 32 (1): 29. DOI: http://doi.org/10.1629/uksg.484

[91] Those who successfully developed standards often cite years to ensure sufficient community consultation and co-development

[92] Sempreviva Anna Maria, Vesth Allan, Bak Christian, Verelst David Robert, Giebel Gregor, Danielsen Hilmar Kjartansson, … Hermans Koen W. (2017, December 12). Taxonomy and metadata for wind energy Research & Development. Zenodo. http://doi.org/10.5281/zenodo.1199489

Recommendation 3: Provide funding for training and for sustaining coordination fora and expert groups to help communities with standard development.

Stakeholders: EOSC, Research funders, Standards bodies, Coordination fora, Institutions
Rationale: It is crucial that communities, especially those less experienced in FAIR practices, have access to people with expertise (for example, data stewards or ontology experts), who can help with development and adoption of standards and methodologies, provide best practice recommendations or case study examples, and offer tailored training. Such efforts have to be appropriately and sustainably funded and research institutions should be encouraged to take long-term responsibility for the availability of such support roles.
Example: Research Data Alliance[93], which already plays an important role by offering a framework for communities who wish to work together, or by offering some excellent tools (e.g. FAIRsharing[94], which is a curated resource on data and metadata standards), or providing recommendations on best practices from various communities[95].

Recommendation 4: Incentivise development of community governance.

Stakeholders: EOSC, Research funders
Rationale: Standards need to be developed by/with the community for them to be accepted and successfully implemented. For this to happen, clear community governance is essential to determine responsibilities and oversight of the different processes and to ensure a structured way of communicating feedback. Such efforts should be incentivised financially (e.g. the costs and time required to organise community consultations).
Example: Astronomy is a discipline with strong community governance. The standard data format for astronomy has been developed in 1981 and endorsed by the International Astronomical Union[96]. The International Virtual Observatory Alliance (IVOA) develops and maintains the technical interoperability standards for astronomy. The IVOA does not have any formal funding, but benefits from in-kind contributions of

[93] https://www.rd-alliance.org/
[94] https://fairsharing.org/
[95] https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs
[96] https://fits.gsfc.nasa.gov/

community members[97], which highlights the importance of advocacy and bottom-up level buy-in for such initiatives to be sustainable.

The wheat research community is an example of a community which used the framework offered by the Research Data Alliance and created a dedicated Wheat Data Interoperability Working Group to facilitate development of best practice standards in a structured manner (clear leadership of the group, clear ways of working and of providing community input, clear timelines and goals)[98].

Recommendation 5: Recognise that FAIR guidelines will require translation for other digital objects and support such efforts.

Stakeholders: EOSC, Research funders, Policymakers, Institutions, Standards bodies
Rationale: Applying FAIR principles to the context of specific communities requires adoption/translation. This need is more obvious in case of other (non-data) digital research objects where a direct mapping of the FAIR guiding principles may not be appropriate. The importance of each principle may depend on the priorities and maturity of the community in their use of certain research objects. This translation will need to be agreed in appropriate community fora, and such efforts should be incentivised financially (e.g. the costs and time required to organise community consultations).
Example: As part of the AGU's 'Make Data Fair' project[99] to enable FAIR Data across the earth and space sciences, town hall meetings[100] and panels[101,102] have addressed the challenges of making other research objects FAIR, including software, samples and workflows. This is beginning to lead to community-specific guidance around metadata and citation practices to improve software and service findability, accessibility and reusability[103].

[97] Genova, F., Arviset, C., Almas, B.M., Bartolo, L., Broeder, D., Law, E. and McMahon, B., 2017. Building a Disciplinary, World-Wide Data Infrastructure. Data Science Journal, 16, p.16. DOI: http://doi.org/10.5334/dsj-2017-016

[98] Dzale Yeumo E, Alaux M, Arnaud E et al. Developing data interoperability using standards: A wheat community use case [version 2; peer review: 2 approved]. F1000Research 2017, 6:1843 (https://doi.org/10.12688/f1000research.12234.2)

[99] Enabling FAIR Data project: https://osf.io/jy4d9/

[100] Data Fair: Sharing Your Software — What Is FAIR?: https://agu.confex.com/agu/fm18/meetingapp.cgi/Session/56228

[101] How Safe and Persistent Is Your Research? https://agu.confex.com/agu/fm17/meetingapp.cgi/Session/25700

[102] FAIR Data Is Not Enough: Communicating Data Quality and Making Analytical Code FAIR I: https://agu.confex.com/agu/fm18/meetingapp.cgi/Session/60523

[103] Hausman, J., Stall, S., Gallagher, J., & Mingfang Wu. (2019). Software and Services Citation Guidelines and Examples. Figshare. https://doi.org/10.6084/M9.FIGSHARE.7640426

Recommendation 6: Reward and recognise improvement of FAIR practice.

Stakeholders: EOSC, Research funders, Policymakers, Institutions
Rationale: Efforts aiming at improvement of community FAIR practices are usually time-consuming and require a lot of dedication. Nevertheless, such efforts tend to be unnoticed in the current academic rewards system, unless linked to journal publications. To incentivise such work and to highlight its importance, it is essential that it is appropriately recognised. This needs to be undertaken by the EOSC, research funders (financial support for such efforts), but, crucially, also by policy makers and research institutions themselves, so that such efforts are taken into account in evaluation, promotion and hiring criteria. This should go beyond merely recognising the efforts of making individual research outputs FAIR. Efforts aimed at greater community engagement, such as development of shared standards for FAIR practices, are crucial and need to be recognised as well.
Examples: There are multiple efforts undertaken by Research funders, Policymakers and Institutions towards better rewarding and recognising researchers for making individual research outputs more FAIR. The final report of the Open Science Policy Platform[104] offers a comprehensive set of recommendations for various stakeholder groups. The Open Research Funders group developed the *Incentivization Blueprint*[105] which provides concrete recommendations with a template specifically for research funders. Initiatives such as the Wellcome Trust's Open Research Fund[106], or the EOSC Co-Creation[107] provide, amongst others, financial support for activities which aim at improving FAIRness of community practices. FAIRsharing is a resource which gathers community standards and credits record maintainers. However, we were not able to identify concrete examples where efforts aiming at improving FAIRness of community practices (thus, at a higher level than just making individual outputs FAIR) were explicitly mentioned in academic rewards and recognition policies. Interestingly, recommendations that such activities should be rewarded have been already articulated in Turning FAIR into Reality Report (Rec. 4, Action 4.1 and Rec. 6, Action 6.2) published in November 2018[108], suggesting that implementation of these recommendations did not happen and should be prioritised.

[104] "Progress on Open Science: Towards a Shared Research Knowledge System" - final report of the Open Science Policy Platform https://doi.org/10.2777/00139

[105] Incentivization Blueprint: http://www.orfg.org/incentivization-blueprint

[106] Examples of projects funded by the Wellcome Trust Open Research Fund: https://wellcome.ac.uk/grant-funding/people-and-projects/grants-awarded?scheme_id=3580

[107] EOSC Co-Creation fund: https://www.eoscsecretariat.eu/funding-opportunities/co-creation-requests

[108] Turning FAIR Into Reality: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

Recommendation 7: Develop and monitor adequate policies for FAIR data and research objects.

Stakeholders: EOSC, Research funders, Policymakers, Publishers, Institutions
Rationale: Policies can be important drivers for FAIR data[109] and other research objects (software, workflows, models, protocols, etc.). Therefore, it is essential that bottom-up, community based efforts are coupled with top-down, policy-driven approaches. Policies should be explicit (e.g. clear roles and responsibilities, FAIR vs open data), aligned with each other, aligned with community practices and other relevant policies and regulations (e.g. research integrity). This applies to policies of research funders, publishers and institutions. Proper implementation, monitoring and suitable incentives are also essential for effectiveness of such policies. Implementation should be coordinated with institutional actors so that demands are not coming into effect without appropriate support and common understanding of means and goals. Western European countries have taken the lead in developing and implementing policies on FAIR. Therefore, dedicated efforts need to be focused on less advanced countries.
Examples: Finnish policies are highly coherent which was achieved through coordination between the developments at a global level (OECD), European level (EOSC and the European Union), national level (Ministry of Education and Culture together with Academy of Finland) and community-level (where both researchers and institutions are present)[110]. National open science working groups[111] comment on policies and ensure that national policy recommendations are taken into account in institutional policies. As a result, the national policy[112] is developed by the community itself (through open science groups), but is at the same time in-line with national and international requirements and funders' demands. The data policy of the Economic and Social Research Council (ESRC)[113] in the UK offers an example of a policy with consequences for non-compliance. It mentions that the ESRC has the right to apply sanctions, such as withholding the final payment of a grant, if data has not been archived within three months of the end of the grant. We were not able to identify published examples of FAIR data policies being thoroughly and transparently monitored.

[109] Science, Digital; Fane, Briony; Ayris, Paul; Hahnel, Mark; Hrynaszkiewicz, Iain; Baynes, Grace; et al. (2019): The State of Open Data Report 2019. Digital Science. Report. https://doi.org/10.6084/m9.figshare.9980783.v1

[110] https://avointiede.fi/en/coordination

[111] https://avointiede.fi/en/open-science-expert-panels/open-data

[112] Declaration for Open Science and Research 2020–2025 https://doi.org/10.23847/isbn.9789525995251

[113] ESRC Data Policy: https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/