Nordic FAIR Data Collaboration Opportunities

NeIC report due Sept 1, 2021, lasted edited Sept 21, 2021

Authors: Andreas O Jaunsen (NeIC), Tuomas J. Alaterä (FSD/TUNI, FI), Philip Buckland (UUmeå, SE), Katrine F. Holmstrand (DTU, DK), Kari Lahti (UHel, FI), Jessica Parland-von Essen (CSC, FI), Benjamin Pfeil (BCCR, NO)

Executive summary

In September 2020 the NeIC Director and board initiated a working group on FAIR collaboration opportunities in the Nordics. This came in the wake of the data tsunami era and the realisation that machine learning and connected data will become of crucial importance to society and science. Furthermore, the findings of the EOSC-Nordic project show that the FAIR uptake in the region is slow and a majority of data repositories are not FAIR, while only a small fraction measures up to the available FAIR metrics. Sharing publicly-funded data in a meaningful, reusable way must become part of the normative behaviour in order to nurture data-driven sciences. By enabling data capabilities we are facilitating interdisciplinary data reuse, along with information and knowledge exchange. A key objective is data integration, where concepts and properties, and the relationships between them are linked, forming knowledge graphs and providing an opportunity to provide insight (new knowledge) from the multitude of relations and patterns.

Although FAIR awareness is growing and there is an appreciation of the Open Science vision among researchers and support staff, a small survey conducted by the working group confirms some challenges that are contradictory to the Open Science (OS) movement and contribute to confusion and status quo when it comes to implementing FAIR. If we are serious about embracing OS and the implementation of the FAIR principles it is necessary to ensure that all elements pull in the same direction (this includes infrastructure, skills and services to policies and incentives). Data stewardship is gradually becoming part of the researcher workflow in some organisations, institutions and departments. Professionalising this crucial support is key to succeeding in a swift and effective transition to sharing discoverable, accessible and reusable data. The working group finds that coordinating tasks such as raising awareness, training and skills development and developing common tools and services for research data management (RDM) can contribute to strengthen the momentum of FAIR uptake and better utilise resources, skills and tools across the Nordics. A Nordic coordination office may be one way to orchestrate a common strategy on working towards the realisation of the Open Science vision.

Introduction

The Nordic e-Infrastructure Collaboration (NeIC) has been effective in stimulating collaboration and knowledge exchange across the region through a number of successful projects executed in the past decade. This effort has facilitated new relations to be created, both within and across science domains, as well as elevating the average competence across these areas through the knowledge exchange that naturally occurs in collaborative efforts to solve similar challenges and reach common goals. This was also the goal when NeIC initiated a series of training courses on 'FAIR data stewardship'. During 2019-2020 about 160 individuals participated in a 40hrs course to develop their skills in FAIR data management – a highly successful and very popular course. Collaboration on data management and skills development holds great potential and executing this within the Nordics in the context of NeIC makes a lot of sense given the aforementioned activities on infrastructure and services development. At the cusp of the transition from the 'data tsunami' (where the focus was on the overwhelming amount of data) to the 'data-driven machine learning' era it is essential that 'Research & Development' departments acknowledge that critical data preparation and curation is needed to drive this new era forward. Organisations and companies that have realised the importance of this are now beginning to develop their data policies, pipelines, training and recruiting data stewards to enable the full potential of their data.

The Nordic countries, with their societal and cultural commonalities, is an excellent platform that has proven to be effective in regional collaborations (cf. NeIC). This has also been demonstrated within the EOSC-Nordic project, where collaboration across borders involves partners from different domains in order to realise a common goal of integrating services and data within the context of the European Open Science Cloud (EOSC). Of particular relevance to the topic of this report is the work package on implementing FAIR in the Nordic region. This work has centered around measuring and monitoring the uptake of FAIR across various data repositories during the 3-year project period to see the effect of activities such as awareness raising, training and direct support. However, there is still more to be done in exploring opportunities for advancing the FAIR ecosystem in a Nordic context, both to strengthen the research community and ensure the broadest possible impact.

The implementation of FAIR is intended to make data discoverable and eventually increase the reuse value of data. Note that FAIR does not imply Open Data, although this should always be a goal – unless there are valid reasons to restrict access to the data itself (descriptive metadata should always be available). Restricting access to data is typically required when it contains person sensitive information or is copyrighted/licensed. Open data and FAIR ties in with the broader vision of Open Science, which according to Foster is defined as follows:

"*Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.*"

A crucial premise for research, therefore, is to ensure transparency of the data, enable the reproducibility of research results, and to build trust in the results and the scientific method. However, a major caveat is that research data is still largely inaccessible, i.e. not shared or not sufficiently discoverable. In the latest report on "The state of Open Data 2020" (Digital Science Report), the longest-running longitudinal survey and analysis on open data with 4,500 respondents, finds that there is still a majority (59%) that think researchers are not getting sufficient credit for sharing data (13% respond it is sufficient). The responses indicate that "*recognition*" is among the more prominent incentives for researchers to share their data (see Figure 1). Policy "*requirements*" and "*public benefit*" seems to also be significant motivating factors for sharing data.

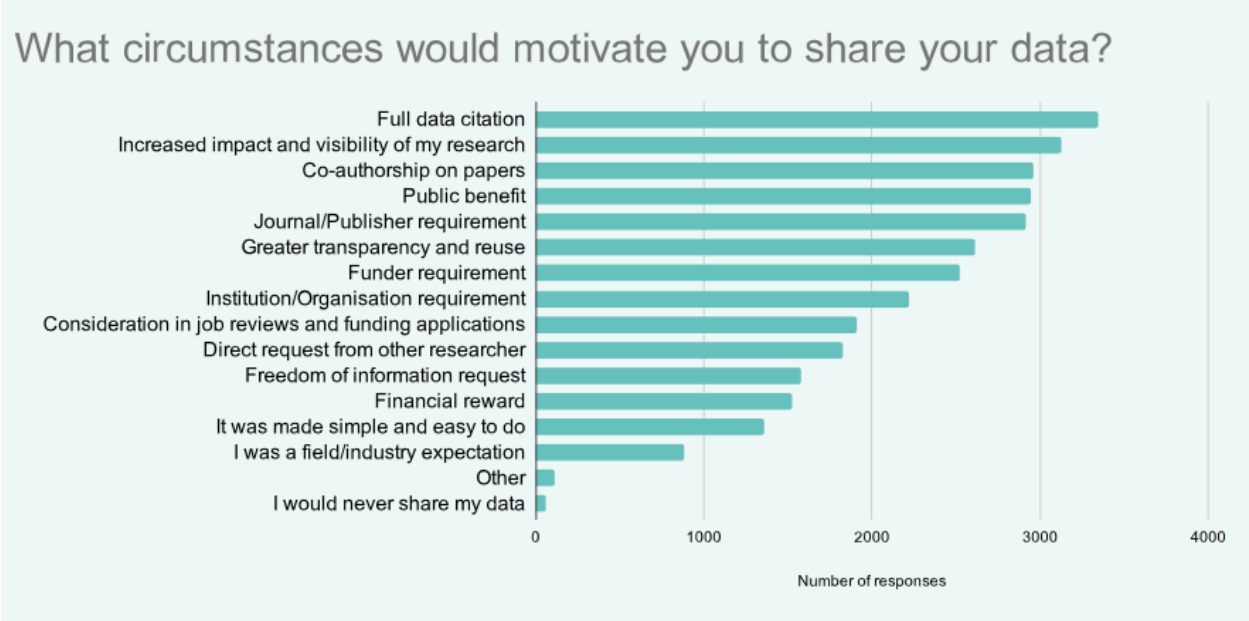


Figure 1: Results from "The State of Open Data 2020" on what circumstances that would motivate researchers to share their data.

The report also indicates that researchers are becoming more familiar with the FAIR principles since the survey began querying this in 2018. However, only 24% of researchers indicate that they are familiar with the principles, which suggests that the majority is still unfamiliar with or even ignorant to FAIR. This is despite the unprecedented attention the FAIR principles (Wilkinson et al. 2016) have received in the 5 years since its initial publication and numerous citations. This is likely related to the social and cultural barriers to change. The existing norms, incentives and policies that are in place within the science community suppress any rapid wide-spread adoption of new practices.

In this report, a working group consisting of researchers and data managers/stewards has explored some of the challenges and opportunities for collaboration on implementing FAIR across the research domains and the Nordic region. The terms of reference requests that the group should provide a science case for Nordic collaboration activities on FAIR research data management in order to best serve Nordic research communities. The report starts with a look

at the perceived current challenges of implementing FAIR in their respective framework. We then present and discuss the areas of focus in two main topics; ‘Policies & Governance’ takes a look at the governance model and how the various elements contribute to the uptake of FAIR and a norm that encourages sharing data. In ‘Data Stewardship’ we presents the skills and career development needed to professionalise the data publication and data management aspects of the research process. In ‘Nordic Convergence’ we provide recommendations that will contribute to a more effective FAIR uptake and discuss why a coherent Nordic approach is beneficial to the region.

FAIR implementation challenges

The survey was an extension of an internal exercise in the working group and answered by 167 respondents in the Nordic region. The respondents were a mix of researchers, data managers and stewards and some miscellaneous academic staff reached via the network of the working group. It was intentionally made anonymous, so the origin of the respondents can not be verified, nor can it be excluded that individuals answered multiple times.

I)	II)	III)	IV)
Which of the following roles describe your position/work tasks (multiple choices allowed)	**Select three (3) challenges you are experiencing in relation to implementing the FAIR principles (Open Science vision) in your work situation.**	**Select three (3) aspects of FAIR that you consider most challenging in your work.**	**Which aspects of FAIR implementation could benefit from a Nordic collaborative effort? (ex: training network, shared expertise, shared ontologies, harmonizing policies / incentives)**

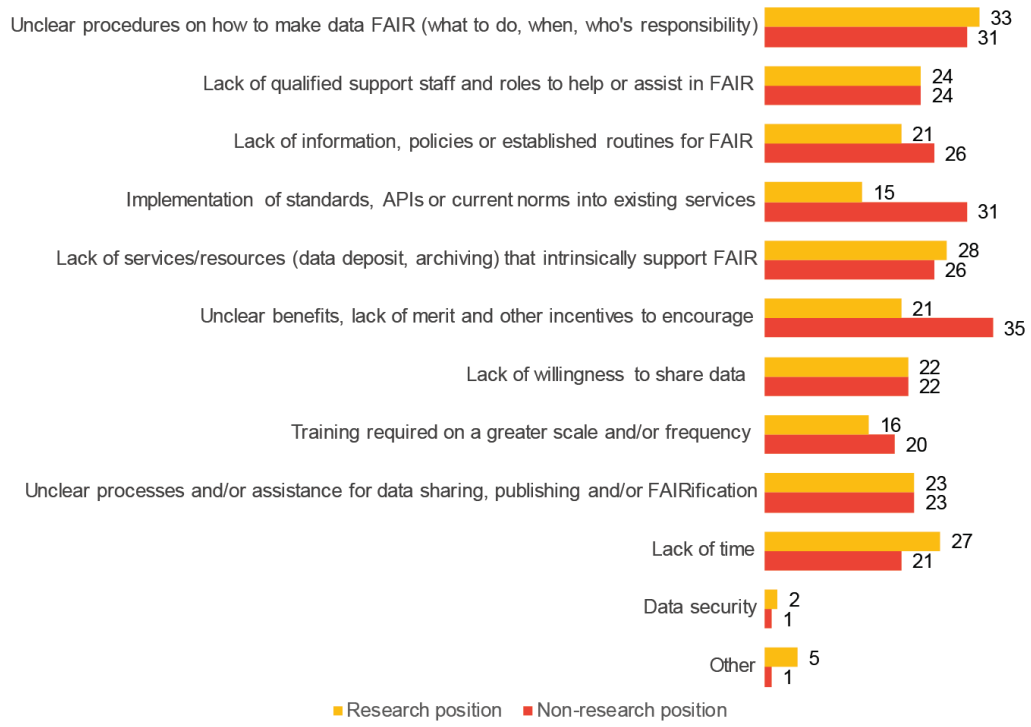
The survey included 4 questions, the initial one (I) to clarify the type of working role of the respondent (up to 3 categories could be selected). The second question (II) was related to challenges experienced in implementing the FAIR principles and the third one (III) on the aspects of the FAIR principles that are most challenging. Finally, an open question (IV) on which aspects of FAIR implementation could benefit from a (Nordic) collaboration.

Results from the mini-survey

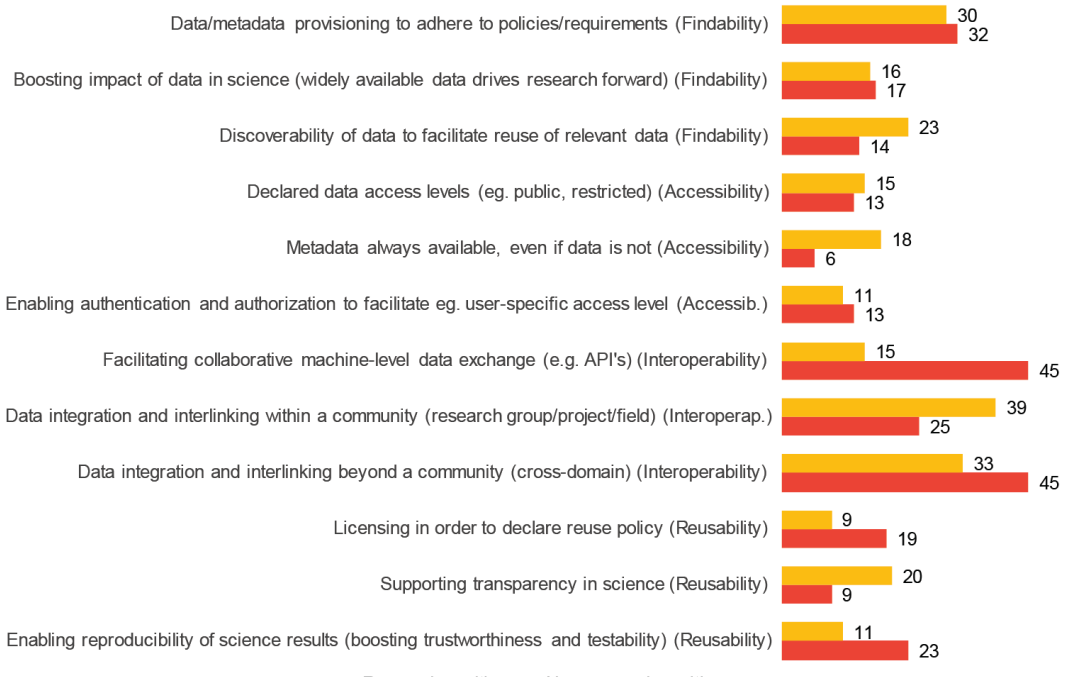
In answering, the respondents often checked multiple working roles. In our analysis, each of these roles was effectively treated as an independent (single) entry with associated answers to questions II, III and IV. Furthermore, the individual role-answer pairs were then categorised into either a *Research position* (researchers, post-docs, students, data scientist) or *Non-research position* (the remaining role categories). The answers to questions II and III were then grouped accordingly and are presented in the following two figures.

For more details on the answers provided, see Appendix A.

II. Challenges experienced in relation to implementing the FAIR principles



III. Aspects of FAIR considered most challenging in one's work



Whilst there is considerable agreement between the answers of the research and non-research roles, there are differences which are worth highlighting. At first glance, issues relating to the implementation of standards and API's appear to be considered more of a challenge for non-research positions. However, this could simply be a reflection of the importance of these aspects for data managers and systems developers tasked with implementing them. That researchers considered this less of a challenge could indicate that they do not consider this their problem or domain, and rely on the other group for implementation. This aspect could also explain the large difference in how much of a challenge 'facilitating collaborative machine-level data exchange' is. Many researchers prefer to download different datasets and integrate them in their desktop software rather than rely on data integration portals. The latter are also few and far between, and tend to either only cover metadata or be related to a particular topic (e.g. https://www.data-arc.org/). Concerns around licensing and data reuse policy may also be similar - many scientists have little concern for or understanding of the specifics of copyright or licensing.

Awareness of the benefits and merit of making data FAIR are, on the other hand perhaps more immediately relevant for those actively undertaking research. Non-research staff still, however, indicated more strongly that lack of clarity in this respect was one of the greatest challenges to implementing FAIR principles, suggesting that it is hard work convincing academics of the benefits of FAIR data. Further work, including interviews with both categories, is probably necessary to investigate the nuances of these results.

In terms of the most challenging aspects of FAIR, there are a number of differences in the survey results. Discoverability of data is considered more of a challenge by researchers, most likely because they actively seek data whilst the non-researchers either provide or manage

data. That metadata is always available, even if data are not, is considered more of a challenge by researchers. It is difficult to know how this challenge was interpreted, but this could reflect either deficiencies in the availability of metadata or deficiencies in the availability of data (i.e. metadata can be found, but the data are not available). The different perspectives and priorities of researchers (within discipline data interoperability), data managers (exposing data from multiple disciplines) may factor into the difference in answers relating to data integration and interlinking. The answers will, of course, be highly dependent on the fields of the researchers, and climate change and archaeological scientists may be more interested in linking data outside of their community than local historians or nuclear physicists. While, interdisciplinary use of data is one use-case that highlights the importance of interoperability, the emphasis on this goal is also crucial for data integration in knowledge systems, where expressing properties and relations across domain silos is essential as a means to enable insight.

The results of the survey will be further used to illustrate points in the discussion below.

Summary of open question on Nordic collaboration benefits on FAIR implementation

The input provided by the respondents in the open question IV can be summarized into the following four categories:

A. FAIR uptake
Rethink communication to researchers and produce clear and jargon-free instructions and guides, both discipline-specific and discipline-agnostic. Create training network(s).

B. FAIR tools and solutions
Offer shared services for software, semantic artefacts, data and metadata that support FAIR.

C. Promoting interoperability on all levels
Agree on shared metadata specifications, vocabularies, identifiers, templates and key standards. Also offer a platform or forum for infrastructures to network. Promote a legal framework to address GDPR and licensing when sharing sensitive data.

D. Collaborate to be part of international developments
Ensure Nordic engagement in EOSC, RDA and https://fairdo.org with a shared agenda.

Policies & Governance

The scope of this chapter is to explore how policies may be invoked in combination with supportive tools to steer or guide the community toward a specific behavior – in this case realising the uptake of FAIR in order to share work and data from publicly funded research. Raising awareness is a key element in advocating the vision of shared data. It is crucial that the

majority of stakeholders have a common and consistent understanding of what this vision is and how it can be realised (e.g. through the adoption of the FAIR principles).

In the context of this document the considerations and recommendations apply to research data that is a result of publicly funded science. The same arguments presumably hold for other types of data, but there may be exceptions that we are not aware of or concerned with here.

The FAIR principles

The FAIR principles (Wilkinson et al. 2016) aim to enable efficient and error-free analysis of data from multiple sources by machines and ultimately by humans, through enhancing their Findability, Accessibility, Interoperability and Reusability. The principles are not only intended for data, as quoted in the original paper: “FAIR differs in that it describes concise, domain-independent, high-level principles that can be applied to a wide range of scholarly outputs”. FAIR metrics and tools have been developed to aid the process of making data FAIR (“FAIRification”). The process requires different types of expertise and should therefore be carried out in either multidisciplinary teams or discipline oriented teams guided by FAIR data steward(s).

To be clear, FAIR does not specify technical details, promote standards or specific solutions. However, when implementing FAIR it is necessary to make implementation choices, typically based on relevance to the community. An implementation profile cannot be copied from one domain and automatically applied to a different discipline. Instead it must be suited for the specific needs and priorities of the community in question.

TRUST and CARE principles in relation to FAIR

When considering a FAIR policy, the relevance of lesser known TRUST[1] principles (for digital repositories) and CARE[2] principles (for indigenous data governance) should be considered.

TRUST principles emphasise the need of repositories to earn the trust of the communities they serve and demonstrate the ability to properly curate the data they hold. A repository certification standard like CoreTrustSeal is an implementation of the TRUST principles that addresses how FAIR data can be kept FAIR over time.

CARE principles address any situation when the data are related to an indigenous culture or about individuals of that culture. In the simplest form, they define that data subjects should have a say and a right to benefit when data about them are gathered and used. The FAIR principles pertain to data findability, accessibility, interoperability, and reusability in a machine-actionable way. They do not directly address ethics or responsibility. However, making rights and governance metadata machine-actionable would benefit both sets of principles.

[1] https://doi.org/10.1038/s41597-020-0486-7
[2] https://doi.org/10.5334/dsj-2020-043

The science case for FAIR

Ground zero for any digitalisation effort is siloed data, meaning data that is generated with no re-usability in mind, often with a single purpose use-case. The data may reside on isolated storage devices and has no or only inadequate metadata. This type of data has a very low potential for shareability and reusability in most cases. Machines cannot self-utilise this data because of the lack of (sufficient) metadata and thus machine learning processes or big data mining efforts cannot utilise this data. Examples of silos are found in departments within hospitals and academic institutions, as well as in many companies. Reasons for the existence and sustenance of data silos are structural (custom software operating on specific datasets), political (proprietorship among data owners), cultural (lack of knowledge and unwillingness to change) and bureaucratic (vendor lock-in).

In order to move away from non-reusable data it is necessary to embrace the Open Science vision and concentrate efforts on how data reuse and discoverability can be maximised. The FAIR principles need to be implemented into the research workflows, tools and repository services, while at the same time being rewarded by funding models and supported by data governance policies. It is important to note that *FAIR data* differs from *Open data* in that the latter promotes openness, while the former highlights accessibility under well defined criteria (data declared as restricted access is compatible with being FAIR). This is mirrored by the popular slogan ‘As Open as Possible, as Closed as Necessary’, used by DANS, EC, GO-FAIR and many others. The goal of FAIR is to maximise the reuse of data, both by machines and humans – supporting the main pillars of Open Science; *transparency, reproducibility and trustworthiness*. It does so by promoting machine-actionable (meta)data discovery, access, interoperability and reuse. The successful and effective implementation of FAIR requires a unison commitment by policy makers, funders, organisations and staff – a level of coordination that is hard to achieve unless it is supported on multiple levels and that the measures are sufficiently synchronised to support the common goals of Open Science. As argued above, FAIR is a major vessel that enables us to move in the direction of these goals. In this respect, the scientific case for FAIR is the vision of Open Science itself.

Transparency is supported by the FAIR principles in that it promotes discoverability, accessibility and reuse of data (sharing the data). Reproducibility is supported by the access to data and any related outputs such as software, tools etc that enables peers to verify the methodology used. Trust is provided principally by the act of sharing the data (and related outputs), enabling transparency and reproducibility. The quality of the shared outputs also contribute to build trust in the researcher or research group.

Culture change

The inertia of “status quo” is substantial. Even if the initial attraction of maintaining the status quo has been overcome and the motivation, skills and knowledge to enable change established – it is not guaranteed that change will occur. This is because changing an individual is not necessarily sufficient to overcome the barrier of “culture”. The cultural and social systems

consist of individuals and these systems shape behaviour by *norms* (what people do or should do), *incentives* (what people are rewarded for doing) and *policies* (what you have to do to be part of the system). Even if one succeeds in nurturing an individual's motivation and ability to change behaviour so that it is consistent with the desired habits, it does not necessarily lead to success due to strong norms, incentives and policies that contradict with the intended behaviour. Changing the culture is substantially more difficult than changing the individual. When norms, incentives and policies are misaligned with desired behaviour, a much more encompassing strategy is needed to drive a successful culture change.

The prevailing practice still is for research teams or individual researchers to operate independently and share their results in publications that summarise their findings. Typically, this means sharing mostly successes, leaving the failed attempts and discarded plans out of the picture – details that could have been valuable lessons to others or provide credibility to the claims. The move to Open Science promotes transparency, self-skepticism, and reproducibility. These traits are largely supported by the research community. So, from an individual perspective everything is ready for a change. However, the research culture in many settings maintains a dysfunctional constellation of incentives and policies that are blocking widespread change toward sharing.

First and foremost, there is the key role that ‘frequent publication in prestigious journals’ plays in securing a job, maintaining it, and earning prestigious scholarships or grants. This culture maintains rewards of publishing findings that are *novel* (as opposed to incrementally building on or affirming existing findings), *positive* (as opposed to negative results) and often contributes to a form of *consensus*. Not everything gets published. The competitive nature of the reward system and the limited career paths in science is contributing to cultural incentives that lead to shortcuts or selective reporting – behaviours that undermine the credibility of the findings (and none of which support showing your work or sharing). Researchers value transparency, but are faced with a contradicting culture that rewards the opposite.

Given these challenges related to the publication culture and incentives, we are unlikely to change the research culture by focusing on individuals’ motivation and ability to be open about their research outputs. Instead, we need a broader and coordinated effort that encompasses the full spectre including adapting services and infrastructures (make it possible), enhance skills and user interfaces (make it easy), change the practice (make it the norm), establish incentives (make it rewarding) and finally policies that pull in the same and intended direction. In short, we need a *change strategy*.

Building on the Center for Open Science’s approach and a lot of similar argumentation, the approach is multi-phased, beginning with the foundation to make it possible; the infrastructure (eg. provide a repository that supports sharing research plans, work and results in accordance to FAIR principles). Next, providing meaningful and efficient user interfaces that makes it easy to share output and integrate with the researchers workflows. By making this practice a norm (eg. starting with students) and integrating with popular tools such as Github, Google Docs etc. The final two levels (incentives and policies), address the reward system and requirements. Rather than maintaining the strong incentives for traditional publications, policy makers can promote

the value for transparency and reproducibility by introducing a system where the importance of the research question and the quality of the methodology is rewarded. Sometimes small adjustments at the right level can have a great effect. Aligning and changing the incentives and policy changes at key decision points and rights stakeholders (funders, journals and research institutions) is seen as crucial to obtain a coherent system that drives and rewards action in the desired direction.

Figure 4: Center for Open Science (COS) multi-level culture change pyramid

A look at legislation and policies

As discussed, it is essential to have a clear policy that rewards the right values and practices. However, the policy alone is not enough, it must be supported by incentives and the entire value chain to make culture change possible; i.e. making it possible, making it easy and making it normative.

Let's look at the current legislation that is relevant in our context.

European legislation on open data

One of the challenges in implementing FAIR data is the lack of knowledge of rules and regulations, legal requirements, and the implications of not following the requirements. With the Open Data and Public Sector Information Directive and the cascading of the regulations in national strategies, we see a path forming for new procedures, funding and in general arguments for implementing FAIR and employing data stewards.

Considering the numerous EU reports that have flagged that the EU needs 250.000 data stewards in academia to help it step into the Digital Age, we also see that pressure from funding agencies to apply FAIR Practices to the entire research lifecycle is increasing, and early adopters have a clear competitive edge when applying to research funds.

There are now a number of EU directives that set requirements to data being ‘open’;

Open Data Directive (Open Data and Public Sector Information Directive)

- Research data are aligned with Public Sector Information (PSI directive).
- Evaluation/definition of the quality of data (high-value datasets) plus longevity.
- Data management planning becomes a standard scientific practice and to support the sharing of research data so that they are findable, accessible, interoperable and re-usable (the FAIR principles).

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 (INSPIRE), Article 17 defines aspects of ‘open’ data.

Regulation (EU) No 511/2014 (Nagoya)

Regulation (EU) 2016/679 (GDPR)

Nordic state of policies on Open Science and FAIR

The Nordic region, with its long tradition of close and effective collaboration, would benefit greatly from coordinating its efforts and policies on Open Science and FAIR. This would enable a well orchestrated approach to adapt to the Open Science vision and allow extensive collaboration among the countries, including skills and resources sharing.

Below we refer to a recent and highly relevant report from EOSC-Nordic WP2 (Deliverable 2.6), where the Nordic policies are studied and summarised in Table 1 (page 28). In the report the authors study both national and institutional level (non-exhaustive) policies covering policies related to the three topics; i) Incentives for FAIR, ii) Open Science training and iii) sharing of software (open source).

Table 1: Summary /analysis

	Incentives for FAIR			OS TRAINING			SOFTWARE		
	YES	NO	DRAFT	YES	NO	DRAFT	YES	NO	DRAFT
Denmark	N			A	N		N		
Estonia		N		A	N			N	
Finland			N	A		N		N	
Latvia		N				N		N	
Lithuania	N			A	N			N	
Norway		N		A	N			N	
Sweden	N			A/M				N	

Note: N= National level policy A= University/institution/service provider level activity M= mentioned in policy/report at national level

Figure 5: Table 1 from EOSC-Nordic WP2 deliverable D2.6 with results from policies analysis on “Incentives for FAIR data”, “Open Science training” and “Software/tools for research”.

For this report we are particularly interested in the ‘Incentives for FAIR’ category, which considers any policy (or directive) that influences or encourages FAIR practices (primarily for data). Their findings (left most, “green” columns) suggest that there are some differences in the progress of policy deployment across the Nordics. In reality, there are no discernible differences in levels of FAIR uptake amongst the countries (cf. private communication, EOSC-Nordic WP4), so if there are differences in policies these have a minor effect or do not (yet) have an impact.

A common Nordic approach on Open Science / FAIR policies?

It is beyond the capacity of this working group to study existing institutional policies across the region when it comes to their mentions of FAIR or Open Science (OS) issues. We have therefore primarily focussed on the national policies that are relevant to OS + FAIR.

Typically, Open Science policies are formulated as recommendations intended to support a vision or bring the practices in line with other concurrent practices. The problem with such policies is that they are not usually given as a directive, but instead are formulated as guidelines or recommendations, which means that they are suggestive and, in practice, optional. This leads to many users selectively adopting them or blatantly ignoring them, because there are no discernable consequences.

Regarding the broad awareness and positive reception with which FAIR has been received from the ground up to high level politicians and meetings (eg. such as the G7 Science & Technology Ministers statements on Open Science in 2017), it is a little puzzling why policy makers are not

more ambitious when it comes to invoking directives that require a certain level of FAIR implementation. Generally, it seems policy makers are hesitant to impose effective policies (directives) due to;

- unclear guidelines for implementation (*skills & experience, incentives*)
- necessary support, tools and services not yet in place to enable the researchers to comply with such policies (*infrastructure*)
- resistance from the community to impose further requirements on researchers (*norms/community*)

These are reflected by the issues presented earlier and listed in the culture change pyramid (see section Culture change to alter predominant habits). This supports the claim made earlier that the problem is related to various aspects of the culture change pyramid. So the current situation is that even though there may be policies in place, the basic foundation for change is largely missing (infrastructure such as support, tools and services) and furthermore with the following layers such as skills/experience, community/norms and incentives. This is also reflected in the mini-survey results, where one of the major challenges seemed to be unclear procedures and lack of guidelines that hinder the implementation of FAIR.

In the next chapter we will discuss a topic that can address a number of these shortcomings and inconsistencies. By professionalising the support and process of data publication (data stewardship) it is possible to address the mid-layers of the pyramid (norms/community, skills & experience). Note, however, that data stewardship needs to be backed up by a good infrastructure on one end and by effective and clear policies and incentives on the other.

Data Stewardship

In the following chapter we take a closer look at a typical data publication workflow from the perspective of a repository. We go on to present the roles of the crucial data stewards that are intended to assist researchers in the planning of data management in a project and take an active role in preparing the data for publishing. Finally, we have a look at some of the skills needed to fill a data steward position.

Data publication workflow

At most institutions and universities an overarching systematic approach for the publication and FAIRification of research data does not exist. Specific archival solutions are in place in various fields of research based upon community agreed requirements in the respective science domain. Examples include the Pangea data repository and the Human Genome Project (HGP). Another reason for domain specific archival solutions is the involvement in European Research Infrastructures which are strongly promoting Europe in the process to develop domain specific FAIRification. This includes tools, domain specific archival solutions, coordination, professional consultancy, access to domain specific data stewards/managers and coordinated workflows in order to make data FAIR and available to the European Open Science Cloud.

The diversity in science is reflected by the existence of domain specific journals where editors, reviewers and typesetters ensure that the scientific and literary work meets the accepted standards of the field. Editors provide guidelines, advice and policies to authors for preparing and submitting manuscripts, establish a system for effective and rapid peer review, make editorial decisions and establish clear guidelines for authors regarding acceptable practices. Reviewers ensure the scientific content meets scientific requirements and typesetters ensure that the format and style meets the set criteria by the journal.

Comparable to scientific publications are the workflow and needs for scientific data publications. Domain specific data stewards/managers fulfil comparable tasks to an editor, reviewer and typesetter and constitute the link between the scientists and the repository. The main difference between domain-specific repositories and general-purpose repositories is that domain-specific repos transfer data as well as metadata into a specific ingest format, ie. data management systems. General-purpose repositories do not normally touch the data (due to the sheer diversity and complexity of dealing with widely differing community standards and practices).

Table 1 describes the typical end-to-end workflow and tasks of a domain specific data steward/manager. It highlights the tasks needed for meeting the requirements of the funding agency (in this case the European Commission under Horizon Europe) that start before the resulting data publication. The basic workflow of the data publication is highlighted in gray. The need for feedback and the complexity of these tasks mean that they can take from between several days up to months. Domain specific tools for communication and submission could speed this process up and experience has shown that first time submissions especially take more time. The development of a systematic archival process, typically as provided by a domain

specific data repository, will solve this problem in the future. It also aims to show what tasks are normally performed by local data stewards that assists scientists at their institution and the tasks of a data manager that works at a repository. Depending on the field and/or institution there can be an overlap and some data managers do tasks of a data steward as well.

Table 1: Typical workflow for scientific data from an external funded project (e.g. by the European Commission). The active data publication workflow is highlighted in gray.

Phase of the project	**Task**	**Responsibility**
Planning phase	Researcher presents data volumes to be expected in an externally funded project	Data steward
Planning phase	Provision of a domain specific strategy that explains in detail how the data/research outputs will be managed in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable).	Data steward
Planning phase	Provision of a data management plan (DMP) that goes into detail for all data generated in the project.	Data steward
Active phase	Researcher opens a data submission process. Researcher includes information on: research field, type of data (observation, model), level of processing (raw, processed, collection of sources), project, time constraints.	Data steward
Active phase	Choice of appropriate archive for the submission (if not already done so in planning phase)	Data steward
Active phase	Provision of a domain specific template for research data in the field, e.g. variable naming convention, units, instrument, calibration, etc	Data manager
Active phase	Researcher provides enriched metadata, title, co-author list and final datasets	
Active phase	Data and metadata are quality checked, harmonized	Data manager
Active phase	Dialogue in case issues occur, e.g. missing metadata or issues in the data itself	Data manager
Active phase	Structuring of all metadata, applies standardized and machine readable vocabularies, reformats data for a machine readable format	Data manager

Active phase	Pre-version (non public version) of the submission is sent to the scientist for approval	Data manager
Active phase	Final publication incl. cross linking with paper publication (if applicable) and PID assignment by the long-term repository	Data manager
Active phase	Ensuring that datasets are properly cited in the paper publication (the same way other paper publications are, including them in the references section). Otherwise, the dataset will not be traceable.	Data manager
Active phase	Feedback about the final dataset publication to European Commission and data management plan is updated	Data steward
Post phase	Ensuring that all data in the project are archived by the end of the project	Data steward
Post phase	Data can be re-used for further research and publications	
Post phase	Data management plan is submitted to the funding agency for final review	Data steward

If we compare the FAIR score (as measured by the F-UJI tool, Devaraju & Huber, 2020) of general-purpose and domain-specific repositories the results are inconclusive at the moment (likely due to a skewed sample). However, apart from some high-profile general-purpose repositories that score well due to high degree of provision of FAIR metadata – the domain-specific repositories generally offer a more advanced service incorporating data processing, analysis and as part of the ingestion process integrate controlled vocabularies, semantics and a generally richer set of metadata. Eventually these repositories will surpass the generalist repositories because they will also incorporate FAIRification of the data content. A dedicated (domain specific) repository, if funded and operated long-term, should be capable of simplifying and providing a highly efficient submission portal with up-to-date standards and guidance provided the service is well-designed. Certification of such services (e.g. CoreTrustSeal) adds an important quality-of-service element that enables users to select suitable and trustworthy services.

Roles of a data steward

Data stewards (DS) are the first point of reference for all data management related questions. They advise, support and train researchers on various aspects of data management throughout the life cycle of a research project, from initial planning to post-publication. This includes storing, reformatting, applying international standards (if possible), managing and sharing research outputs such as data, images, models and code.

Data stewards also advise researchers on the ethical, policy and legal considerations during data collection, processing and dissemination. In a way, they are general practitioners for research data management and can usually solve most data related problems faced by academics. In cases that require specialist intervention, they also serve as a key point for referral (eg: IT, patent, legal experts).

Data stewards may be ideally suited to implement the research culture change described above. They can help raise awareness of the legal requirements and need for sustained data management, good science practices and the benefits and impacts of such practices. Each and every faculty / department will typically benefit from dedicated data stewards. The tasks include, for instance, implementation of research data policy, assisting in writing DMPs, integration of data management for proposal writing, training of researchers and students (workshops), information sessions, networking to specialists such as semantic experts, data scientists, legal advisors or data modelers. Data stewards should also ensure data management plans are reviewed and followed up, and that their data are Findable, Accessible, Interoperable and Reusable as required by the host organization or funding agencies. This requires knowledge of the organisation's policies, available tools, guidelines and a knowledge of the available resources within a given domain or discipline. We envisage that an organisation-wide team of data stewards would consist of both generic stewards, who are familiar with the relevant policies, resources and skilled personnel – and of domain experts (embedded data stewards) who are knowledgeable about the prevailing standards, vocabularies and semantics of a science field. Both types of support staff would be involved in the preparation and publication of research data.

In summary, the researcher brings the knowledge and experience of the domain and current best practices, whereas the data steward has the technical skills to create and use metadata vocabularies and templates. In the creation of high quality FAIR data and services, both expertises are absolutely essential, and the researcher and data stewards form a team.

What skills are needed for DS?

Data FAIRification requires different types of expertise and should therefore be carried out in a multidisciplinary team guided by FAIR data steward(s). Here are some of the areas a data steward is expected to either have elementary competence on or, in case of domain specialists, have detailed knowledge of;

1. An overview of national and institutional data policies and understanding of how to meet the requirements therein
2. General understanding of the scientific process and how researchers work and interact with supporting staff
3. Overview of typical data formats for the major science domains in the organisation
4. Familiar with the FAIR principles, Open Science vision and stay on top of developments in this area
5. Familiar with FAIR metrics and experience with most relevant FAIR assessment tools

6. Familiar with common generic metadata standards, relevant persistent identifiers (PIDs) and other identification schemas for persons, organisations, publications etc.
7. A basic knowledge of semantics & ontologies and familiar the most common controlled vocabularies
8. Knowledge of available resources for storing and publishing data and metadata (general purpose and domain specific repositories)
9. Data modelling and conceptual modelling
10. Experience with publishing open linked data, knowledge of science impact parameters and CRIS systems

A solid long-term approach to creating a data steward competence center is to organize a team that covers most of these skills, including staff that has domain specific knowledge spanning the larger science domains in the organisation. It is also useful to have a network with access to skills that reach beyond the team in the rare occasions where this expertise is needed. The core of such a team may be formed by data stewards, with knowledge of the national and institutional policies and requirements, available resources including relevant data repositories that are likely to be used and a solid knowledge of the FAIRification process in general.

Early adopters

The Netherlands has displayed a remarkable and ambitious adoption of the data steward role on a national level. A recent national report on Professionalising data stewardship in the Netherlands illustrates this. Vocal ambassadors that have helped put FAIR on the agenda in the past years include Barend Mons, who stated “It is irresponsible to support research but not data stewardship” in a Nature article and recommends investing 5% of research funds in ensuring data are reusable.

The Technical University of Delft (TUDelft) was one of the first research institutions to introduce data stewards rooted in the University library. It was part of the Dutch roadmap towards national implementation of FAIR data stewardship - a strategy for professionalizing data stewardship in the Netherlands.

The library formed a core group as a RDM competence centre for training and education of data stewards, and with the three other Technical Universities they formed the *4TU Centre for Research data.*[3]

Https://www.openaire.eu/blogs/setting-up-a-data-stewardship-programme-an-institutional-perspective.

In short, every faculty of the 4TU has its own Data policy based on The Dutch National Programme on Open Science. The local faculty policies also include implementing faculty Data Stewards coordinated and trained from a central RDM centre at TU Delft library, but each with

[3] 4TU.ResearchData is run by the 4TU.ResearchData Consortium which consists of Delft University of Technology (TU Delft), Eindhoven University of Technology and University of Twente. 4TU.ResearchData is hosted and managed by the TU Delft Library.

discipline specific knowledge. A data Steward must be familiar with the research fields standards and vocabulary.

For the researchers the Data Stewards are the first contact points for any data or software questions. If the DS cannot answer, the DS can refer to a specialist (e.g. Legal, Library, ICT and more) The Data Steward role advances around three practices: 1. Active data handling (RDM), 2. Governance and protection of data and 3. Creating the foundation for collaboration.

Basic job profile components of a data steward – as illustrated in the report p 40 *Professionalizing data stewardship in the Netherlands: competences, training and education* 10.5281/zenodo.4320504

This is undoubtedly an example of a public organisation (Research and Education) that is an early adopter of organisation-wide data stewardship. It is still early days, but this hints at the transition that is needed to drive the large-scale implementation of Open Science and FAIR data. In recent years we have begun to see examples of this appearing in smaller scale (isolated departments or research institutes) and in some commercial companies that are data-driven. In the appendix we provide one case example of a Nordic early adopter utilising data stewards in their work.

Nordic convergence

In the preceding chapters we have highlighted major obstacles that need to be tackled in order to realise the Open Science vision and specifically to implement FAIR. It is also important to keep in mind that when the data has been prepared for reuse it must be maintained by curators to remain reusable. In this chapter we present recommendations and areas of possible collaboration in order to succeed in converging toward mature data products and metadata that maximises the capabilities for discoverability and reuse of research data.

The opposite of data management is not ‘no data management’, but rather ‘bad data management’. The basis for convergence on FAIR is to ensure good data management throughout the research workflow and by facilitating domain specific accommodations. By joining forces and harmonising the process in the Nordics we can benefit from skills exchange and join forces in areas such as training, access to specialised skills and use of common tools. As we have seen in the chapter on Policies & Governance it is critical to ensure that researchers and support staff experience that infrastructures, services, expertise, incentives and policies pull in the same and desired direction. Currently this is not the case and a culture change is needed to orchestrate a coherent move towards Open Science values and implementation of FAIR.

Facilitating convergence on Open Science

FAIR is not an objective in itself, but rather a means to an end to facilitate Open Science. The sharing of FAIR data and software should be seen as a part of the research process and the workflow should be made “FAIR by Design”. In section Culture change we discussed why it is crucial to adapt the infrastructure, provide user-friendly, effective interfaces, develop skills and expertise that can help change the norm in communities, aided by incentives that rewards such behaviour and to make certain actions required through policies.

Infrastructure

The adaptation and support for FAIR in the infrastructure layer is crucial to simply make it possible for users to abide by policies or expected behaviours. This support needs to be worked into both generic and domain-specific data repositories. Typically, we would expect domain-specific repositories to gradually excel in its support for FAIR and easy data submission procedures due to their ability and knowledge of the specific needs, standards and practices within a given domain area. A relevant obstacle will be how to determine what repositories to support in developing their FAIRness. The reality of the matter is that in certain domains there are numerous repositories already available, while in others there may be few or none. The currently available repositories support FAIR in varying degrees. So how to determine what repositories to develop further and what mechanisms could be used to stimulate creation of repositories where there currently is none?

The German research council (DFG) has taken an apparently successful approach to this. The DFG initiated in 2019 its NFDI programme that is funding research data management and the implementation of FAIR in national consortia with an annual budget of €70M. In this way, communities need to self-organise to create consortia that collaborate to apply for funding from the programme. Successful applicants get funding over 3-5 years to mature their FAIR support within the domain / community.

Interfaces and tools

For example, the FAIR implementation requirements should be reflected in Data Management Plans (DMPs) – a formal document that states how data will be handled during and after the execution of a research project. DMPs are increasingly becoming required elements that are part of funding proposals. However, their use and quality are seldom questioned. It is important that funders and researchers alike understand why DMPs are needed, how to write them, and ultimately how to use them. Crucially, DMPs should be machine-actionable, meaning it must be possible to automatically check that commitments made in the plan are followed through by the project – e.g. by assessing the FAIRness of published products (datasets).

Skills and expertise

To offer services that support FAIR (or provide FAIR as a Service) requires considerable resources, a majority of which are human resources that support and guide researchers with tasks ranging from data analysis, data management, data curation, data storage, publication and repository service development. Making and keeping data FAIR is a joint effort that starts with the researcher who generates the data, involving in varying degrees the aforementioned roles and tasks. The repositories that house the data are key resources that are expected to specialise the FAIR aspects for the relevant data using insight, standards and new technology. The data knowledge and content is addressed by domain and semantic experts. Keeping data FAIR long term cannot rely on the researcher, nor is it a static piece of documentation. The data and metadata must be curated to take into account new standards, controlled vocabularies and domain practices. Preserving the comprehensibility of a dataset requires domain specific knowledge and continuously updated data management guidelines, enabling the long-term data curation. A FAIR policy should identify the goals for competence building which involves continued investments in data management skills, infrastructure and data support. For the FAIR and open science ecosystem it is vital that the data stewards have training but also meaningful career paths and communities/networks of expertise.

Data stewards are just now becoming important additions to support staff in specific research teams and in a few larger research organisations. We believe these resources will be essential in realising the visions of Open Science. In section “What skills are needed for DS” we provide more details on what a data steward is and what skills they should have.

Incentives

Several Nordic research councils have for some years now requested researchers to plan their data handling via data management plans and generally recommend abiding by the FAIR principles. Examples include the Swedish Research Council, Norwegian Research Council, Academy of Finland.

Mons (2020) recommends that 5% of all research funds should be spent on research data management (RDM) in order to ensure that data are made available and reusable. Sadly, this is considered to be highly controversial by various stakeholders and there is generally little determination to realise the Open Science vision through a solid and effectful investment. Furthermore, it is still acceptable practice that tens of millions of euros are invested in projects or infrastructure that generates data, without any realistic plan or budget for how those data should be managed and published. This is an area in which funders hold a strong card with the power to impose a potentially highly effective incentive, and one that would hopefully be backed by national policies. By requesting publication of the science outputs within a reasonable time limit, registering science outputs (using DOIs or similar PIDs or GUPRI[4]s) that can be verified through machine-actionable metadata content, funders can effectively track the output and 'openness' of projects.

Policies

The assumption is that the intention of policy makers is to implement the majority of elements of Open Science in order to support basic principles of science (transparency, reproducibility and trustworthiness) and maximise the value and impact of the investments into science (reuse of data and tools, constructive feedback on plans and preliminary results/analysis, enable validation of research results and peer-review publication). As argued in a previous section, policies must be synchronised with incentives, the support apparatus, user interfaces and resources to reflect the intentions of the policy makers.

An excellent example of an ambitious and clear policy is the French national policies on Open Science, Ouvir la Science.

List of recommendations

Although we can hope for continued small (uncoordinated) steps and uptake across the mentioned areas and that this eventually turns the tide of current norms and common practices on sharing of data and software in research – it is rather unlikely to be successful by the end of the decade. However, one would hope that a more coordinated approach would be established across the region to ensure a faster and more effective adoption of Open Science. What would such a coordinated effort consist of provided there is political willingness to implement the OS vision? We propose some concrete actions divided across a few themes:

[4] Globally Unique and Persistent Resolvable Identifier

a) **<u>Share research output</u>. In order to enable publication of outputs it is necessary to develop and maintain a pool of data stewards. It is necessary to establish and integrate a force of capable data experts (e.g. data stewards, data managers, repository operators, curators etc) across the Nordics. The skill sets must include a profound understanding of the full eco-system of FAIR data, including knowledge of metadata standards, semantic interoperability, digital objects, provenance, licensing and more. Recommendations include:**

   - **Set up training program for data stewardship and advanced career development that can apply to a broad group of career paths including data stewards, data managers, repository developers and curators, researchers and support staff**
   - **Establish university level training on data stewardship and data management for researchers, support and curators. Furthermore, secure alignment of curriculums for data stewardship across the Nordics and and exchange programme for data stewards**
   - **Establish a Nordic train-the-trainers network for scaling up and evolving the FAIR Data Stewardship practices. A key goal here is to achieve a homogenous skillset and practices for the data stewardship role in order to stimulate mobility and common practice across the region.**
   - **Continue hosting Nordic courses on FAIR Data Stewardship, developed and executed by members of the aforementioned trainers network (e.g. coordinated by NeIC). The course can be adapted to suit specific communities and will require repetition and should be supported by practical exercises.**
   - **Hosting and coordinating workshops is a step in the direction of more advanced training and community development. Examples of this can be BYOD (bring-your-own-data) workshops or M4M (metadata-4-machines) events, where specific needs, relevant standards, vocabularies and ontologies are identified or created.**

b) **Beyond the skills development effort it is essential to get institutions onboard to professionally develop the data publication by creating organisation-wide data stewardship competence centres. This will help professionalize the service and operation.**

c) **Provide sustainable infrastructure to support the publication of FAIR data, software and other research outputs.**

   - **Targeted and sustainable funding of excellent domain-specific repositories that hold a potential to support the domain on a national and/or Nordic level. Such repositories not only provide the requirements and streamline the publication of data for the relevant domain, but also develop and adopt**

standards of relevance, implement controlled-vocabularies, and develop tools to provide effective and quality controlled data submission procedures for the community. In some cases these repositories may collaborate in larger international networks on behalf of the domain or at the very least unite efforts (and possibly the service itself) on the Nordic-level for a cost effective and coherent implementation across the region.

- **Provide generic data repositories to meet the data and outputs publication requirements of science areas and data types that are not covered by dedicated domain-specific repositories. Generally, it will not be possible to achieve a similar level of FAIR compliance in such repositories as the it will not be possible to adapt the submission, formatting, standardisation to the same degree as in a given domain/community.**
- **Ensure that any supported repository abides by a minimal level of FAIRness and that it is adequately certified (e.g. CoreTrustSeal)**
- **Harmonisation of data management and services to make it easy for users to share 'FAIR data'. As far as possible, generation of data should be 'FAIR by design', meaning the reusability aspect of the data must be kept in mind when designing and preparing data generating pipelines. Making research data FAIR is not the same as keeping them FAIR. Data curation is an integral part of any long-term data archive and certification is an important aspect of securing that such services abide by best practices and can be trusted with the commitment implicit in providing data long-term.**

d) Develop mechanisms and incentives that stimulate sharing and reuse of data by rewarding a behaviour that is characterised by Open Science practices

- **Implement the requirement to disseminate and share publicly funded research data, preferably along side provisioning or support for developing data repositories that make this process as easy as possible**
- **Facilitate easy access to and discovery of FAIR data. This could be done via a common data repository that provides customised solutions to the various domains (or separate repositories for different communities). The provisioning of an infrastructure removes a major obstacle that prevents users to share their data (and to safeguard it for the future)**
- **Encourage best practices in data citation, eg. by providing recommendations or guidelines on how to use persistent resolvable identifiers for individuals (ORCIDs), science outputs (DOIs), research org. Registry (ROR) and more.**
- **Facilitate the tracking and impact of scientific outputs in order that funders and authorities can verify adherence to policies and requirements. This would essentially enable a form of Open Science Barometer, where a researcher's 'openness' and productivity could be gauged**

- **Minimise publication bias (the tendency to only publish positive results and to not publish negative or inconclusive studies) by diversifying the science impact from being exclusively directed towards sensacional results to incorporate diversity, broadly acknowledge other science outputs such as software and data, in addition to (some form of results publication)**
- **Encourage and facilitate the reproducibility aspect of science, growing the trust aspect to scientific results and outputs. The practice of selectively considering one's own data in a study should be turned to regularly incorporate comparative analysis of available FAIR data that is relevant for the topic**

Nordic coordination office for FAIR and Open Science

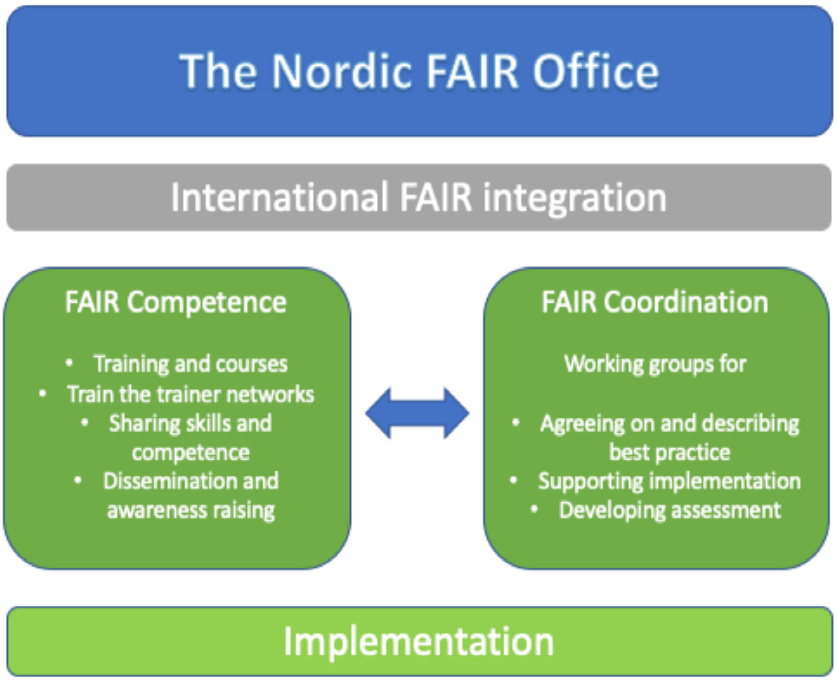
As mentioned earlier there is no one-size fits all solution to the implementation of FAIR and chances are it would likely be rejected by many communities. A community-based (domain) approach is more likely to succeed as it is necessary to adapt the implementation and solutions to the various communities. Still, there are ways in which we can reach better interoperability and develop robust solutions for individual communities.

One intriguing idea is to create a Nordic shared vocabulary or ontology for FAIR and Open Science related topics. Expanding along this line of thought, and in light of the areas identified in Culture change, we would improve from sustained and determined efforts to turn the somewhat confusing and counterproductive scene into a coherent forceful push for culture change (challenges related to infrastructure, skills & interfaces, norms, incentives and policies). We believe that prolonged (long-term) support and stimulation in these areas would provide a lush environment that could permanently change behaviours and norms to a more open and peer-feedback based culture.

A Nordic FAIR Office could coordinate initiatives and activities across the region over the course of a 5-10 year period (building to a large degree on the GO-FAIR initiative). We have identified four themes that could be taken as examples of working groups active in conjunction with the FAIR office under the coordination function. The aim is not to duplicate activities that are done elsewhere, in EOSC, RDA or similar projects, but to support their uptake as development is very fast, as well as to facilitate community needs to be communicated to the international implementation and standardisation work. An expert body like a Nordic FAIR office can also support national policy and competence development.

To support the implementation of the FAIR principles in the Nordics and promotion of the Open Science vision would be greatly benefitted by a regional coordinating office to ensure solutions that are viable, in line with international initiatives as well as serve the research communities. We propose creating a Nordic FAIR Office to initiate and execute activities that support

competence building and coordination of core activities – all supporting the implementation of the FAIR principles across countries, domains and multiple stakeholders.



The office would have an overall coordinating role of activities related to Open Science and in particular to FAIR data (and other artefacts). It would also coordinate with international stakeholders and communicate global aspects of relevance on the topic of FAIR (calls, best practices, technology choices etc). Internally, in the Nordics, the coordination would primarily be towards the national providers that also facilitate infrastructures for data-driven science.

Obtaining FAIR convergence requires both top-down (e.g. government, funding agency, university policies and requirements) and bottom-up (researcher and data user driven) approaches. Information needs to flow from the experts in different fields to the policy and coordination activities, but there is also a need for goal oriented work towards alignment by creating shared practices and interoperable solutions. This is dubbed ‘FAIR Coordination’ in the illustration. Training and awareness raising, or ‘FAIR Competence’, are equally important and should be closely linked to the Coordination activities. The Nordic countries can also coordinate their efforts in the international projects and networks, like EOSC and RDA, so that information can be shared actively and good Nordic practices promoted in international contexts.

The creation of a Nordic FAIR office focusing on both competence building and coordination of development could support key activities and strengthen the Nordic FAIR capability by producing both training and developing, refining and FAIR implementation in alignment with international efforts.

Recommended areas for centralised support

A. FAIR implementation and interoperability support

We should further develop a shared understanding of what FAIR and research data management entails. This can be done through creating a shared terminology in the Nordic languages based on existing ontologies in the field, a Nordic shared vocabulary or ontology for FAIR and Open Science.

There could therefore be efforts done towards a vision of a linked universe of knowledge, that could be exemplified by

a. Different ontologies in domains (e.g. https://bioportal.bioontology.org/)
b. Nanopublications and knowledge graphs (e.g http://www.kulttuurisampo.fi/?lang=en)
c. Translational tools (e.g. https://perio.do/en/) and interdisciplinary research (e.g. https://www.data-arc.org/)

Furthermore, the implementation of FAIR can be supported by common **recommendations and solutions for pan-Nordic generic descriptive (discovery) metadata and semantics** could be produced and published as a central web resource (Nordic FAIR specifications) where there is need to complement international infrastructures and projects. These should be based on and aligned with European and global recommendations and could consist of

- Registry of commended shared international semantic artefacts (per domain?)
- Registry of commended descriptive metadata elements
- Recommendation for expressing access and restrictions
- Recommendation for expressing contributor roles and the CRediT taxonomy
- Recommendation on identifier use
- Recommendations for expressing these in JSON-LD? Create schema/SCHAL expressions ?
- Developing guidance on FAIRification of software and scripts. Legal aspects of data ‘ownership’ and licensing of all types of outputs

B. Strengthened data management planning

Good data management planning is of essence for successful alignment with the FAIR data principles. Collaboration within the Nordics on data management planning tools and services would be beneficial. A first, low hanging fruit, could be a common Nordic-wide template, alternatively a set of templates, for data management plans (or within the Data Stewardship Wizard). Such templates should also be aligned with EC and other relevant stakeholders. Further on down the line, could be the development or implementation of support for an

interoperable machine-actionable Data Management Plan, i.e. DMPs that are FAIR data objects and that autonomously connect to FAIR research outputs. Ensuring at a minimum level Nordic DMP interoperability would be beneficial (e.g. FI: https://www.dmptuuli.fi/, DK: https://dmponline.deic.dk/) and the guidelines presented by Science Europe form a practical framework through which to aim for convergence. This could be achieved by collaborative production of the following managed documents and activities:

- Recommendations for DMP format (data model, metadata, semantic artefacts)
- Enhancing machine actionability (e.g. DSW and similar solutions)
- Linking DMPs to other tools like CRIS, service providers and service catalogues
- Finding sustainable business models for DMP services
- Creating recommendations for APIs in DMP services (open programmatically access to DMPs for evaluation and as research data)
- Solving related legal, copyright and GDPR issues (templates for common issues)
- Engaging funders (standardizing the requirement for, and evaluation of, DMPs)

C. Strong shared training and developing structures, knowledge sharing and ways of work that are cross-border

There should be shared Nordic networks in place for sharing knowledge, organising common training and developing services and interoperability. Networks, common projects and different kinds of collaborative spaces should be offered to data stewards, service providers and researchers. Also, ensuring **sustainable solutions and expert resources** bridging over projects is important. For this organisational engagement is necessary. Training resources should be shared and coordinated over the Nordics.

- Research oriented advice on achieving goals legally and ethically (e.g. with GDPR)
- Advice on appropriate data storage and transfer tools for different legal frameworks
- Developing a pan-Nordic data service infrastructure that fulfils all relevant national legal obligations
- Manage clashes between national interpretations of laws and desirable international collaboration, also contrasts between corporate and state (university, government etc.) possibilities
- Providing domain specific support for legal interoperability by creating a shared understanding of e.g. GDPR implementation
- How to manage sensitive data
- Creating awareness of data needs and solutions between nations *and* disciplines
- Nordic graduate courses on data management

D. Clear and sound incentives and merits for FAIR alignment

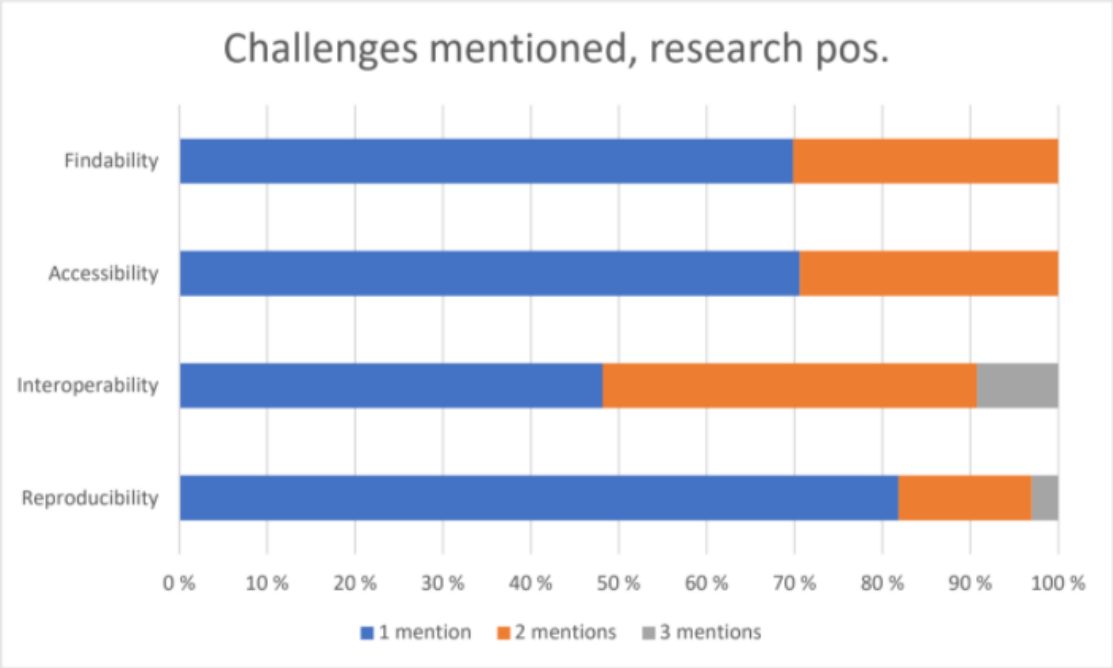
Concerning legal and organisational interoperability there are still things to do, despite the shared culture. The incentives differ between communities and stakeholders. **Data policies should be aligned with FAIR and offering the researchers support and credit for open and FAIR publication of research should be promoted.**

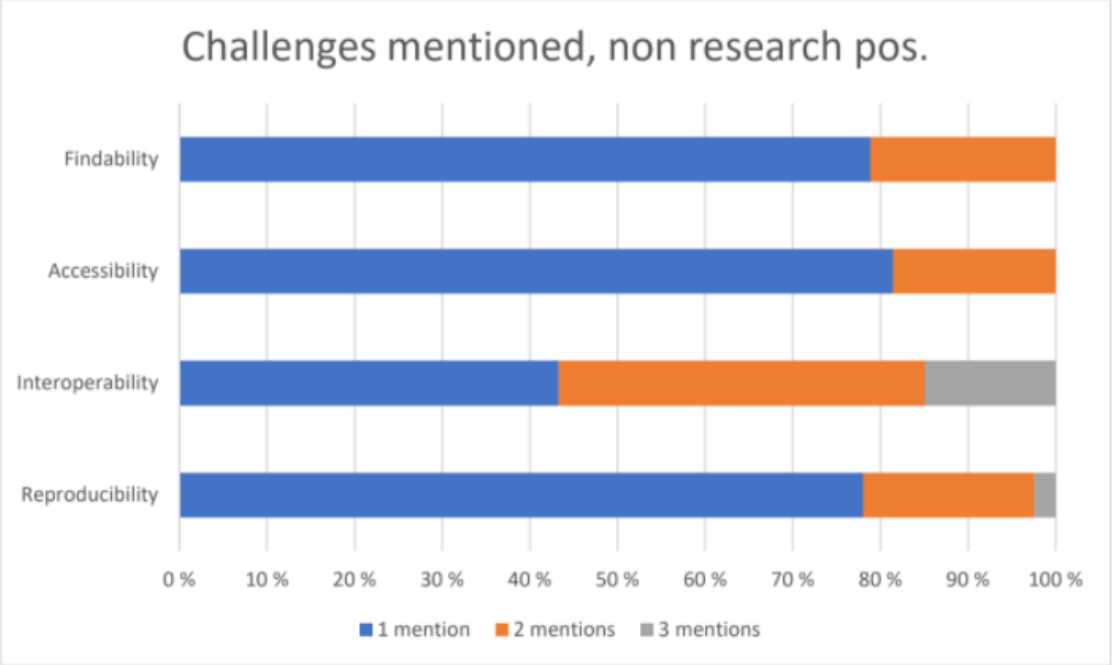
- Harmonize data policies in the Nordics so as to ensure that these are aligned with the FAIR principles
- Ensure that data policies are in place
- Develop common Nordic merit systems for data sharing and publication
- The EOSC-Nordic project has a task and associated deliverable on Nordic FAIR incentives
- Recommending use of DataCite Contributor Types https://schema.datacite.org/

Appendix

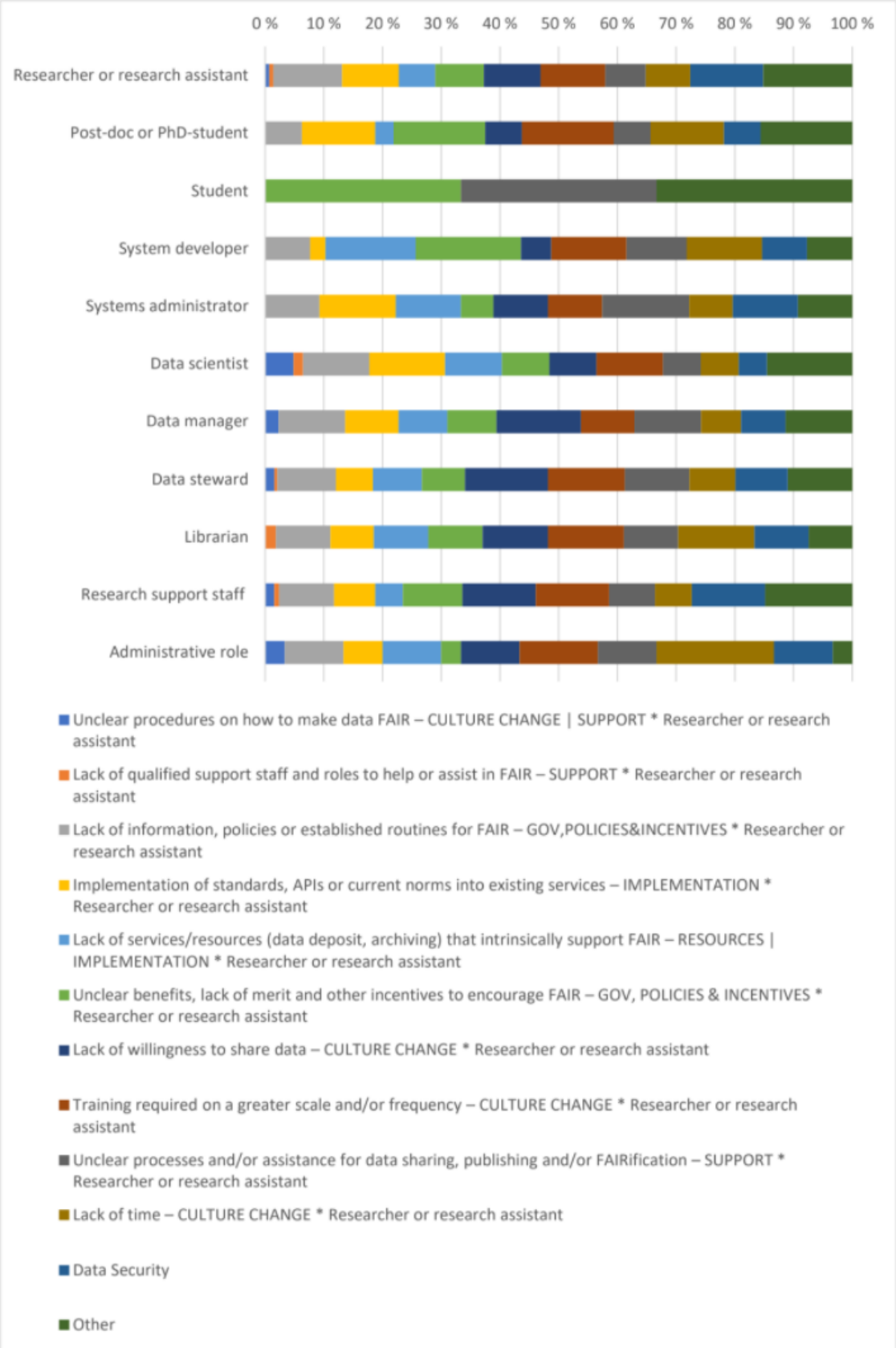
Results from the mini-survey

Question 3, answers grouped by F A I and R. Interoperability seems to be the most challenging one.

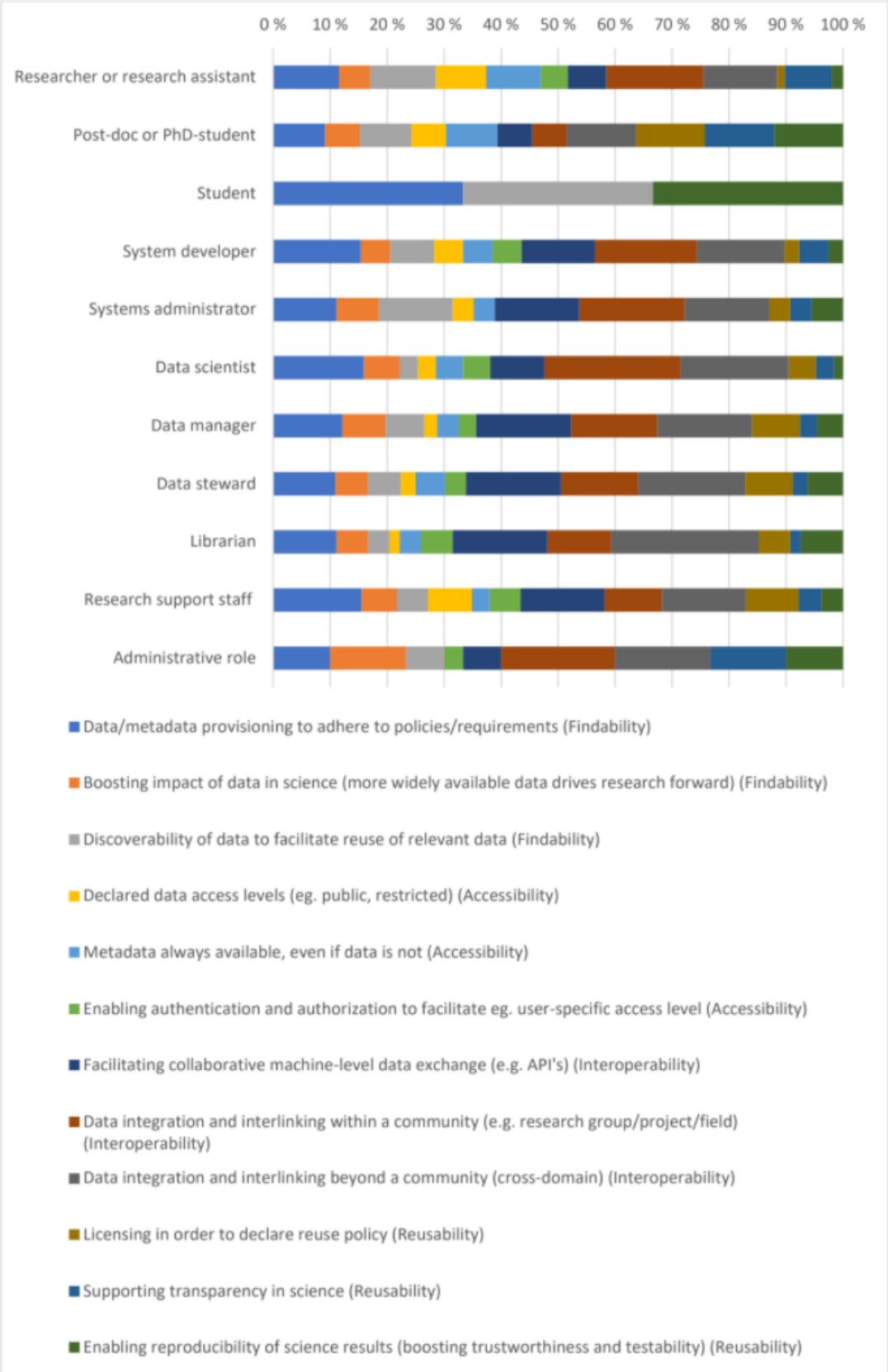




Challenges experienced in relation to implementing the FAIR principles. All mentions by the position/work role (%)



Aspects of FAIR considered most challenging in one's work. All mentions by the position/work role (%)



Comments on ‘research data requirements’ from legal advisor, Erica Schweder (GU)

Is research data normally considered public data or does the legislation distinguish between them?

Yes, research data that either enters or is compiled at a state university is in most cases judged as public documents. In certain individual situations, research can be counted as "case processing" and this has the effect that research data becomes public documents in connection with the case being closed and taken care of for archiving. That research is counted as a case is quite unusual, so the starting point is that research data is to be regarded as public documents. Should anyone request research data, they can then be disclosed if they are not covered by confidentiality.

Are there any requirements in Sweden, the Nordic countries or the EU for research data to be shared, and in that case there are some requirements for> how the data is to be shared (metadata, metadata + data, machine readability etc.)

No, there is no legal requirement to share your research data. On the other hand, funders and journals where researchers must publish articles most often require that research data be made available. This can be done in different ways depending on the requirements of the financier or the magazine. It is often referred to that data is made available subject to national rules on confidentiality / confidentiality. Some journals require that research data be available in special repositories, often it is then metadata that is made available.

Is it primarily the responsibility of institutions, research funders or authorities to introduce guidelines / requirements for> researchers about and how data should be shared?

The responsibility for having guidelines for how research data is to be made available is probably shared between universities, authorities and funders in my opinion. As long as it is not a legal requirement to make research available, there is really no obligation to have any guidelines. On the other hand, the requirement increases because many people think that research data should be made available anyway, even though it is not a legal requirement.

It is important to distinguish between open data and the principle of openness here - even if you do not make your data available, someone can always request data according to the principle of openness. However, this can lead to parts not being handed out due to confidentiality.