National strategy for data management based on the FAIR principles

Preface

Research creates results that can be utilized by society to address both global and national challenges. The ongoing digitalization means that research results can be shared and used easier than ever before. Research data can be used and reused by other researchers, and results can be reproduced for comparison of results. This promotes more effective and better research, and it increases confidence in the research conducted.

The European Commission has launched a number of policy and strategic initiatives aimed at supporting better research and societal use of research data, including the European Open Science Cloud, the object of which is to establish a Web of FAIR data and services to facilitate finding and reusing research data. In 2018, the Danish Agency for Science and Higher Education published a Strategy for National Cooperation on Digital Research Infrastructure. This data management strategy, based on the FAIR principles, is an important initiative in the 2018 strategy and should be seen as a first Danish national response to the requirements of the European Open Data Directive the (PSI Directive) that national policies must be drawn up for making research data accessible in accordance with the FAIR principles.

In order to preserve research data and make them available to others, it is necessary to keep track of the contents and physical location of the individual data sets throughout the research process. This is known as data management. To ensure that data can be found, accessed, interoperated, and reused, they must comply with the FAIR (*F*indable, *A*ccessible, *I*nteroperable and *R*eusable) principles. The purpose of this strategy is to ensure that research data generated via public funds, and possibly co-funded by private research-funding foundations, meet the FAIR principles.

The target group for this strategy is researchers and management at Danish universities, preservation institutions which support research, and other institutions that conduct research using public grants. The strategy lays out a number of principles and identifies actions and initiatives aimed at advancing the process towards a successful implementation of the strategy, including the development of practices for how research data are handled.

The strategy includes many different research areas and institutions. An open approach is therefore needed where not everyone is expected to be able to use the same solution and not everyone can proceed at the same speed. This will require resources, which must be found in an interaction between the institutions and the research-funding foundations.

This strategy has been prepared by DeiC on assignment from the Danish Agency for Science and Higher Education, and it is the result of a wide collaboration, including contributions from many relevant stakeholders and individuals. More than 700 comments and proposals for the draft strategy were received during the consultation process alone. We are looking forward to following the next steps towards a national implementation of data management based on the FAIR principles at Danish research institutions.

Director General Hans Müller Pedersen   Professor John Renner Hansen

Contents

Introduction

1.1 Purpose and vision

The purpose of the strategy is to establish a basis for decisions on national implementation and financing of a data management practice based on the FAIR principles[1] i.e. the data must be Findable, Accessible, Interoperable[2] and Reusable, and the conditions for use of data must have been clearly formulated. The strategy must be in accordance with the EU's Open Data Directive[3] and the Danish implementation of this in the PSI Act[4], and contribute to meet research and societal expectations for increasing accessibility of research results.

The strategy thus supports the directive's requirement for the greatest possible public accessibility of publicly funded data in the field of research. In addition to increasing the possibilities of reuse, compliance with the principles in the strategy can contribute to higher reproducibility of research and thus to compliance with the *Danish Code of Conduct for Research Integrity[5].*

The strategy focuses primarily on 'FAIR' rather than on 'open' data, as data can be FAIR without being open, i.e. freely accessible. Data management in accordance with the FAIR principles is thus also relevant in relation to data which cannot be made openly accessible for ethical, legal, commercial, or other reasons. However, under this strategy, FAIR should be understood as meaning 'as open as possible, as closed as necessary', i.e. data are to be made open, unless there are explicit reasons against this.

FAIR data can, in themselves, be seen as a process rather than as an either-or with the possibility of compliance with several or a few of the principles. The purpose of the strategy is thus to contribute to taking significant steps towards making research data more FAIR and to research data being treated and recognized as a resource shared in national and global data infrastructures in ways that stimulate open research and open innovation.

The overall vision for research data management is that researchers and other players continuously make qualified decisions and actions in relation to maintenance of and access to their data (and any other results/products such as software, protocols, or the like) with due consideration for the promotion of research, the research institutions' policies, existing agreements, the needs of society, and the existing legislation. The intention of the strategy is to contribute to making it easier for researchers to share and reuse data, leading to more FAIR data of high quality and thus to better research opportunities.

The strategy must make recommendations on how to finance the handling of research data in accordance with the principles established and to make proposals for policies for data management in accordance with

---

[1] Wilkinson, M.D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci.Data 3:160018 doi:10.1038/2016.18 (2016)

[2] The term 'interoperability' is used in the FAIR terminology as well as in this strategy.

[3] Directive (EU) 2019/1024 of the European Parliament and of the Council on open data and the re-use of public sector information (2019)

[4] The current *Danish Act on re-use of public sector information (Lov om videreanvendelse af den offentlige sektors informationer)* is being revised to implement the latest version of the Directive.

[5] Danish Ministry of Higher Education and Science, https://ufm.dk/publikationer/2015/filer/file (2014)

the FAIR principles for projects funded by the research-funding organizations. Overall, it can be said that the strategy objective is to establish principles and the accompanying actions for a FAIR data management practice and its funding, as well as actions for the realization of the strategy.

Although the FAIR principles of 2016 are relatively new, they can in a sense be understood as a summing up and structuring of already established practices. Digitally strong research environments, such as the European ESFRI projects[6], have organized themselves and built solutions that largely meet the goals and principles of the strategy for a number of years. For academic and professional groups that already use primarily discipline-specific international infrastructure solutions or established institutional solutions, this strategy will probably not entail any significant changes; they will still be able to use existing infrastructures and research support organizations.

For other academic fields that may be data-bound to a lesser extent or that have not organized themselves in international infrastructures to the same extent, the strategy may offer a change. In this context, the individual academic fields play an important role in defining how and to what extent they will implement the FAIR principles in their research practices. The implementation of the strategy is expected to stretch over a number of years, with an increasing share of FAIR data in line with the provision of the support structures.

The strategy does not constitute an exhaustive treatment of all relevant FAIR data issues, but sticks to some key tracks that can help establish and disseminate a data management practice that supports the FAIR principles. More details may be included in a coming implementation plan. However, in practice, opportunities and restrictions will undoubtedly emerge, and they must be exploited and addressed, respectively. In the implementation of the strategy, it will be up to all the parties involved to support and collaborate on this in the coming years.

1.2   Background

Increasingly, society and research are using various types of data as a basis for identification, analyses and modelling, and it is therefore important to create opportunities to find, access, interoperate and reuse data. The plan to prepare a new national data management strategy to support that more research data become FAIR originates from the Strategy for National Cooperation on Digital Research Infrastructure[7].

The strategy also serves as the Danish accessibility policy under the EU's Open Data Directive (formerly the PSI Directive), which, in its latest version, also comprises research data and requires that the member states draw up policies for research data accessibility in accordance with the FAIR principles[8].

The objective is to make data as open as possible in accordance with the FAIR principles, so that data can be accessed and re-used by others to the greatest possible extent. However, it should be stressed that FAIR data are not the same as open data. There may be good reasons why data cannot be made openly

---

[6] European Strategy Forum on Research Infrastructures, see https://www.esfri.eu/

[7] Strategy for National Cooperation on Digital Research Infrastructure, Danish Agency for Higher Education and Science   (2018)

[8] Directive (EU) 2019/1024 of the European Parliament and of the Council on open data and the re-use of public sector information (2019)

accessible. The Directive thus mentions that issues relating to intellectual property rights, the protection of personal data and privacy, security and legitimate commercial interests must be taken into account. However, data that cannot be accessed openly can still often be made FAIR, as the researcher may publish metadata describing the actual existence of the dataset, as well as conditions for how others may be able to access the data, i.e. either the dataset in its entirety or derived versions thereof.

Denmark is a member of European Open Science Cloud (EOSC)[9], and the expectation is that, in future, research data can be made accessible through EOSC. At the time of writing, an infrastructure for this is not yet in place, and the EOSC architecture is still being developed. However, FAIR plays an important role in the visions on EOSC, and meeting the FAIR principles will in itself be a step towards making data accessible through EOSC. To some extent, the strategy also leans on the recommendations from the EOSC FAIR Working Group, which focuses on the establishment of FAIR practices[10]. In relation to the establishment of infrastructures and services, these should generally be based on well-known standards and on well-known requirements for architecture and interoperability in EOSC[11].

It follows from the explanatory notes to the draft for a revised PSI Act that the EU's General Data Protection Regulation and the Danish Data Protection Act (*Databeskyttelsesloven*) will always take precedence over the provisions in the Open Data Directive and the PSI Act[12]. However, even personally identifiable data can often be made FAIR, even if they are not made openly accessible. In addition, the data management activities covered by this strategy may involve a number of issues regulated by other legislation and local policies. These include the Danish Archives Act (*Arkivloven*), security policies, etc. Such matters must obviously be taken into account in the practical data management activities, even if they are generally not addressed in this strategy.

1.3   Scope

In accordance with Article 10 of the Open Data Directive and the Bill for the Danish PSI Act, the present strategy applies to publicly funded research data, meaning the result of (or output from) publicly funded research and research conducted by public institutions where there may be private funding as well. In accordance with this legal basis, the strategy is aimed at research institutions and research-funding organizations. In particular, the Directive stresses that research data which have been made accessible by researchers, research institutions, or research-funding foundations through an institutional or thematic repository must, to the extent possible, be reusable for both commercial and non-commercial purposes[13].

---

[9] https://eosc.eu/

[10] See https://eoscsecretariat.eu/eosc-fair-wg-outputs

[11] See, for example, https://eoscsecretariat.eu/eosc-architecture-wg-outputs

[12] It is recognized that the further details on clarification of the legal basis may pose difficulties in international collaborations between countries with different interpretations of GDPR, copyright and the like.

[13] The Open Data Directive, Article 10(2).

Research data are defined in the Open Data Directive[14] as *documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results.*

The strategy uses the term *data*, understood as the digital content of any kind that can be referenced in the form of data sets, files and databases, etc. This may include all types of digital research output that form part of the research. Be they logbooks, software, workflows, and scripts, etc., according to the EOSC report *Turning FAIR into reality[15]*, which instead of 'data' uses the term *FAIR digital objects* to clarify the breadth of the research output.

The focus of the above definition is that data must have been collected or produced in connection with research activities regardless of the organization thereof or the further circumstances connected with these. For example, data collected or produced in connection with scientific advice will consequently be covered as long as the scientific advice is based on research. Regarding data from research-supporting collections such as libraries, museums, and registers, the validity of this strategy is limited to the data which are made a direct object of the research, as stated in the above definition[16]. Notwithstanding this, increased accessibility of such collections in accordance with the FAIR principles could certainly contribute significantly to the research conducted in Denmark.

The focus of the strategy is on the handling of newly produced research data, and it thus does not aim at a comprehensive retrospective process. This obviously does not exclude that there may already be even highly valuable data with individual researchers or institutions that would benefit from being handled in accordance with the principles of the strategy.

1.4 The value of open and FAIR data

It is a fundamental principle of the sciences that researchers build on each other's results. In a world in which research is becoming increasingly digital in line with other activities in society, not only publications, but also data, must be made easily accessible, so that it is possible for researchers to incorporate other researchers' data into their own research. A larger data basis can provide new opportunities for analyses and models, for example climate models, analyses of large text corpora, etc. Good data management practices based on the principles of this strategy can contribute to making this possible in practice. FAIR data must benefit both research and society as a whole. For some fields of research, this is already an established practice, where, for example, computational power and data infrastructures across countries and organizations are a prerequisite for being able to make a mark in the field in question. This practice is expected to spread to more research fields, and the objective of the strategy is to contribute to meeting the researchers' needs.

---

[14] The Open Data Directive, Article 2(9)

[15] *Turning FAIR into reality. European Commission*, DOI:10.2777/1524 (2018)

[16] This does consequently not address any obligations of the institutions in question under other articles and sections of the Open Data Directive and the PSI Act.

FAIR data will also make research outputs more accessible to society in general, facilitate the research and development work of companies, and be of significance to citizens' performance of citizen science activities. In addition, making the data available to the extent possible will contribute to transparency and thus to research credibility.

The prerequisite for making these visions realizable and to benefit research is that data are findable and identifiable, and that there must be relevant, meaningful, and adequate metadata connected, including that the conditions for reuse are stated in the form of an attached licence. These are some of the objectives of the FAIR principles, which are fully designed to enable re-use of data in a responsible manner, for the maximum benefit of research and society in general.

1.5 Target group and structure

The strategy outlines a number of principles for data management practices that support FAIR data, followed by recommendations for actions aimed at strengthening the implementation of the individual principles. An attempt has been made to structure these actions in accordance with the key players in the implementation thereof. The recommended actions and initiatives often support more than one principle, but are listed here under the area in which they play the significant role.

The principles are broadly aimed at all stakeholders in such practice, including researchers, data stewards, research managers, and research-funding foundations. The proposals for action areas are, however, primarily aimed at decision-makers in research-conducting and research-funding institutions. This means people who can assume responsibility for the implementation and funding of the strategy, including how the work is organized and carried out, as well as for the establishment of the necessary policies and organization within the individual organizations.

The principles and action areas are not prioritized, as there is an implied interdependence between them.

As an actual implementation plan is expected under the terms of reference, the strategy does not comprise a proposal for specific implementation, but only proposals for *action areas*.

2 Principles for a data management practice that supports FAIR

2.1 Data management must support the FAIR principles and opportunities for data sharing and re-use.

1. When data are assessed by researchers or a research institution as potentially being of value to other parties than the researchers themselves, it must be considered how and to what extent data can be made accessible.

2. As a general rule, metadata which describe data can be made open, also in cases in which data are not openly accessible.

3. Data are secured a minimum level in relation to the FAIR principles by being uniquely identified by an identifier (*Persistent Identifier*, PID) and described with metadata.

4. Compliance with the FAIR principles is a step-by-step process in which the individual FAIR principles can be met gradually.

2.2 The relevance and weighting of the individual FAIR principles vary between research areas and must be defined within these

1. The level for compliance with FAIR, in addition to the minimum requirements described above, is defined by the individual research field in relation to international norms and standards for the field.

2. Researchers must work towards consensus within their respective research fields on the level of FAIRness that they are aiming for, including which FAIR principles they consider the most important.

3. The individual research projects lay out the process for achieving the desired level of FAIRness as part of their data management plan.

2.3 It must be possible to preserve all types of research output and file formats of relevance to the research in the short term and the long term

1. Data that are regarded by researchers or a research institution as having a potential future value must be stored and made accessible in a technically and organizationally secure data infrastructure in accordance with the FAIR principles.

2. In cases where, for example for financial reasons, it is necessary to delete data that have already been made accessible, particularly valuable data must be identified and secured long-term preservation.

3. In cases in which data that have been made accessible are deleted, the PID should remain accessible and be able to generate metadata about the deleted data as well as the information that data are now inaccessible.

2.4 Data management is a continuous process throughout the course of the research project

1. Research projects plan and document data management during the course of the research project in a data management plan that is continuously maintained.

2. Data management methods and tools must be accessible to help researchers collect and store data and metadata during the course of the research project.

3. The conditions (licensing) for reuse of data must be made accessible as part of metadata, in both human and machine-readable form.

2.5 Relevant technical infrastructure and tools must be offered to all researchers, regardless of academic, professional, and institutional affiliation

1. Infrastructure for data storage and accessibility in accordance with the FAIR principles must be available to all researchers employed at Danish research institutions.

2. The technical infrastructure must provide the best possible support in meeting the research institutions' obligations in relation to the Danish National Archives' demands to avoid double registration.

3. The technical tools and platforms chosen, be they commercial or open source, must make it possible to extract data and make them FAIR in an open, standardized form.

2.6 The necessary help and knowledge resources must be accessible regardless of academic, professional, and institutional affiliation

1. It must be possible for researchers to work with their research data based on the FAIR principles without having to be data management experts. If needed researchers must be offered sufficient and competent help from their institutions.

2. Competences in data management and FAIR principles are developed in a process involving researchers and support functions.

3. Data stewards must be available at several competence levels and with sufficient expertise in the corresponding research disciplines.

3 Actions for the establishment of a data management practice that supports the FAIR principles

3.1 FAIR data management is a change of practice in many academic fields

A research world in which researchers not only share their methods and data, but also make them accessible in ways that make the data reusable, will help change the perception of what is useful and usable research output. This has implications not only for the researchers' working methods, but also for a number of the surrounding structures, including:

- For the merit transfer system, which must recognize new areas of work in connection with this change in practice

- For the grants system, which must recognize that making metadata and data reusable is time-consuming and thus entails an additional cost

- For the recruitment system, which must recognize the need for new competences.

This will require action by the research institutions, which will contribute to a change in practice in several academic fields and work actively in relation to the metrics used in evaluations and appointments.

Research councils and foundations will need to incorporate not only support for this type of activity, but also recognize actions and initiatives in FAIR data management as credit-bearing activities.

Research institutions should stimulate and facilitate this change in practice, particularly in academic fields which do not currently have a tradition for sharing data outside a closed group, including:

- That the research strategic importance of good data management is recognized. Research in an increasing number of sciences is today entirely dependent on data being collected, systematized, categorized, analysed (e.g. by computers), stored (e.g. in repositories or databases), and re-accessed by international research colleagues.

- To facilitate an understanding that the FAIR principles do not necessarily mean open access to data, but that also closed data can be made FAIR by being, as a minimum, linked to an identifier (PID) as well as descriptive metadata.

Research institutions should, in collaboration with the Danish Ministry of Higher Education and Science, work nationally and internationally with the credit transfer system so that researchers' work with publishing, maintaining, and supporting the sharing of FAIR data is recognized and rewarded.

Proposals for specific actions:

- Strengthening of communication and dialogue on FAIR data management.

- Start-up of a process focused on making more research outputs credit-bearing and ensuring recognition of data management with a view to FAIR. This must be done at a national level with the involvement of international initiatives. The credit transfer should take into account differences in research areas.

3.2 Expert groups are supported in defining their own implementation of FAIR principles

The individual academic fields have different traditions for how data are shared and structured, and data are included in different ways and to different extents, taking into account the methodological approaches and established practices of the discipline in question. It is therefore important that the researchers in the academic environments are those who define how and how far FAIR can be implemented in their fields. The steps that are taken towards the provision of FAIR data must be based on good scientific practice in order for the result to be of value to others.

The importance and structure of data can vary widely both within and between academic fields. In some cases, this is about personal data, in others it concerns data that may have been created in a private-public partnership , which can be subject to confidentiality agreements, or data may be copyrighted, to name some examples. In addition, there may be ethical reasons why data cannot be published, for example in the form of a special confidentiality relationship between researcher and informant. However, many research data can be made open without major legal and ethical challenges.

Research groups should prepare a plan for their support of FAIR data. Such a plan may, for example, determine which metadata to expose and which data types, ontologies, standards, PID systems, etc. that are relevant to use for the research in question.

Research institutions must support the work of the academic fields to prepare plans and, taking cost into consideration, plan the principles with which they will work and how far they will get with FAIR within the next three years through their support functions.

Research institutions and DeiC[17] must collaborate on ensuring that the right competences and tools are available to support the researchers in formulation and definition of their FAIR plans.

Research institutions must support researchers in requiring that data published as part of publication in scientific journals or other types of publications must be made FAIR to avoid double publication.

Proposals for specific actions:

The research institutions must ensure that descriptions of good data management practices are prepared in different academic fields, preferably based on international practice. As a minimum, these must contain a description of how metadata can be collected as an integral part of the work process. This may entail:

---

[17] DeiC has been established as a collaboration between the eight Danish universities and the Danish Ministry of Higher Education and Science. The basis for DeiC has been described in '*A National Strategy for Research e-Infrastructure*'.

- Preparation of short-term goals for making data accessible and sharing data, for example based on methods regarding maturity models[18].

- Helping expert groups identify, develop, and/or publish standards, methods, tools, and infrastructure to make their data FAIR which are based on, or at least interact with, international solutions.

- Working with data management practices for a selection of data and with methods for quality assurance of data and metadata.

- Preparing principles for correct (re-)use of data.

3.3 Supporting preservation of research data in both the short term and the long term

Preservation of data after a project ends requires both human and technical resources. The best way to preserve data is to keep them 'alive', thus ensuring that they are relevant and continuously accessible, and are used in accordance with the research standards applicable at any given time. This requires stable funding and organization and is a prerequisite for the ability to meet the FAIR ambitions.

The primary aim of this strategy is to make research data accessible in accordance with the FAIR principles. In some cases, this may be limited to making data available while a research project is active, but it will often involve retaining and curating data after the research project has been concluded, typically in an infrastructure or data system specifically designed for this purpose. In the Open Data Directive, such data systems are called institutional or subject-based repositories.

Preservation for 5-10 years, corresponding to, for example, the requirements of the *Danish Code of Conduct for Research Integrity*, will, as a general rule, consist in the operation of a repository with the necessary level of governance, information security, and documentation to ensure confidence in the contents as research output. In some, but not all, cases, this will involve certification in accordance with an international standard. The aim should be to create, in as far as possible, an international setup for systems in which data are shared, for example under the auspices of international research infrastructures. In addition to technical infrastructure and IT systems, the operation of data repositories also requires data-literate staffing and therefore stable funding.

It may be difficult to determine what will potentially be valuable to preserve for posterity, and there are several known cases in which unnoticed research outputs or publications have taken on new significance many years later. In many contexts, it will nevertheless still be necessary to choose what to preserve in both the short term and the long term. The considerations may be connected with limitations in available resources of a human, economic, financial, and technical nature.

The selection must necessarily be based on the present conditions and criteria. At least the initial selection should be made by the researchers themselves, primarily in accordance with criteria regarding expected

---

[18] For example CMMI Institute's *Data Management Maturity (DMM)*, (https://cmmiinstitute.com/data-management-maturity) or RDA's FAIR Data Maturity Model: https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

research and societal relevance. The selection criteria may be influenced by policies for the repositories or archives used for continued preservation and accessibility. Selection procedures and criteria can be incorporated in a data management plan for the given research project.

In some cases, storage in excess of a 5-10 year period may require special measures, and there will also be a need for expertise and stable organization over an extended period of time. Over time, more stringent requirements will also presumably be made for data to remain FAIR and understandable outside the original framework of understanding. The general rule should be that data are stored and made accessible in durable, open formats where the specification of the format is well documented and publicly accessible.

Where, after consultation with the researchers, the Danish National Archives receives data for long-term preservation under the Danish Archives Act and makes these data accessible in accordance with the principles laid out in this strategy, the requirements of the strategy can still be said to have been met.

The recommendation in the first five years after the strategy has entered into force is to focus on short-term preservation in order to take significant steps to preserve data on a large scale and generally make such data accessible. The issue of specific requirements for infrastructure, organization, and funding for long-term preservation should be re-addressed before 2025. However, the question of the preservation period will depend on academic fields, and storage for longer periods may already now be part of the policy for some discipline-specific repositories.

Research institutions should work with the various academic and professional groups to map existing stable repository solutions and identify where there are no such solutions. In addition, principles and methods must be developed to decide which data to store and make available as well as guidelines and criteria for when data have lost their relevance and can be erased.

Research institutions must contribute to the establishment of a practice and resources for continuous data management, also after the research has been concluded. This includes dealing with questions about responsibility for the continued management and accessibility of data after researchers have left the institution where data have been produced.

Research institutions and DeiC must collaborate on ensuring the operation of institutional and national data warehouses and repositories, which must have sufficient long-term funding and continuous ongoing protection and curating of data, including identification of data for long-term storage.

The Danish National Archives must ensure that the research data they receive for archiving are made accessible in accordance with the FAIR principles, in as far as possible.

Proposals for specific actions:

- Collaboration, also with international players, on the development of methods and infrastructure for short-term and long-term preservation of research data adapted to research requirements for diversity in research output and file types, as well as scaling in relation to data volumes.

- Collaboration on defining criteria for the value of data in terms of reusability and long-term preservation.

3.4 Support research requirements for documentation throughout the research process

All information, structures, metadata, etc. necessary for data and other research outputs to be sharable with others should be documented as an integral part of the research process. This can be done by ensuring systematic collection, which is described in a data management plan updated throughout the research process. An increasing number of research institutions and funding organizations require that a data management plan must be prepared, for example in connection with applications for funds. Good data management practice helps ensure research transparency.

At the end of a research process, there should be follow-up on the data management plan, where it is reconsidered which data and research output must be preserved, where, for how long and according to what licences, including whether there are data-supporting scripts and methods that can be advantageously shared with others.

Funding organizations should enter into a dialogue with the research institutions to examine how they can support and stimulate through their policies that data (and other outputs) are made FAIR.

Research institutions must support researchers in building up well-functioning procedures for continuous collection of metadata based on existing good documentation practices in the different disciplines.

Research institutions must ensure that a clear division of responsibilities is established in relation to all aspects of data management in compliance with FAIR and actively support data management practices. Dialogue and communication with researchers should underpin an overall picture of good data management practices, including, for example, information security, GDPR, licences, accessibility, long-term preservation, etc.

Research institutions and DeiC must facilitate the widespread use of data licences.

Proposals for specific actions:

- Recommendations must be prepared within the academic fields for methods to ensure continuous documentation of research data in line with the data being collected and generated, so that all necessary information is present when results are to be made accessible.

- A discussion on risk management in connection with FAIR data should be stimulated and facilitated at national level. Examples of risk factors are confidentiality, copyright, other intellectual property law regulation, dual use, GDPR and national security interests.

- Institutional guidelines for the use of data licences must be drawn up and disseminated, possibly in national cooperation, primarily based on international standards.

3.5 Establishment of technical infrastructure that supports data management and accessibility of FAIR data

Researchers at universities and other public research institutions must have access to the necessary technical infrastructure for collection, storage, and accessibility of data. What is relevant will depend on the project, the size of the project, and the level of internationalization. As research is often inter-institutional and international, this infrastructure should be seen as part of a global ecosystem, including EOSC. Regardless of whether the infrastructure is created locally by the institutions, as joint national solutions or as part of international collaborations, it is important that the solutions offered are anchored in the needs of the research in question and how best to meet them.

Infrastructure solutions must respect the international dimension of the research and must therefore comply with international standards. Where recognized international solutions already exist, for example open-source software, such implementation solutions should be preferred over own-developed solutions. This applies to the implementation of infrastructure and technical tools, as well as to the choice of metadata standards, vocabularies, etc.

The infrastructure must provide the best possible support for the researchers in avoiding double work in connection with notification and possible submission to the Danish National Archives. To begin with, the aim is especially on supporting workflows in, for example, repository systems, so that the Danish National Archives is advised about additions of new data and then itself retrieves the metadata needed for the data to be regarded as having been correctly registered. In the longer term, workflows between the research infrastructure and the Danish National Archives should be completely automated, so that double work can be avoided completely in connection with both notification and submission.

The research institutions must ensure the establishment and maintenance of the basic infrastructure that is made available to all the institution's researchers, including data storage, networks, etc., corresponding to current needs.

The research institutions must collaborate on identifying and making proposals for the parts of the technical infrastructure which can be solved nationally, taking into account the international development and national and local needs and experiences.

The research institutions, the Danish National Archives and DeiC must work together to establish workflows that can result in the registration of necessary metadata in the institutions' data systems being accepted by the Danish National Archives as compliance with the duty of notification.

The research institutions, the Danish National Archives and DeiC must collaborate on ensuring, in the longer term and no later than in connection with the revision of the Notification Order in 2025, that there is cohesion and interoperability between their respective technical infrastructures, so that processes can be made fully automated.

Proposals for specific actions:

Specifying and constructing national parts of the technical infrastructure:

- Establishing a generally accessible nationally trusted data repository where researchers can store and expose metadata and data, as well as attach persistent identifiers (PID) and licence.

- Specifying and offering components nationally that are demanded by the researchers, for example PID systems, ontology databases and other components that can facilitate the work with making data FAIR.

- Extending the national service for data management plans with academically customized templates as well as adaptations that can make data management plans usable in connection with registration in the Danish National Archives.

- Ensuring that recommended services and infrastructures are certified in accordance with international standards as required, and that governance and policies underpin correct management of access to data even after researchers have left the institution.

3.6 Establishing a knowledge infrastructure that provides help and knowledge resources for data management and FAIR

The changes towards research becoming more data-driven and more international affect research methods and require new expertise and specialized support functions in addition to new technical tools. These traditions are already established in many fields of research. For others, it will involve a big step from recognizing the importance of data management to becoming an active part of infrastructures and continually creating the necessary metadata. There have been calls for the establishment of a new profession, 'data stewards', who contribute to facilitating data management processes, and to making and keeping data FAIR.

'Data stewards' with in-depth knowledge of data management issues, methods, law, and technical tools must be able to enter into research projects in line with, for example, laboratory technicians, programmers, engineers, or librarians. This field of competence can be found in variants ranging from very general competence at a more advisory level, to subject-specific competence at the highest academic level, which is fully integrated in the research.

The construction and organization of data stewardship competences which can form part of close collaboration with the researchers are an essential prerequisite for ensuring that FAIR data do not entail a significant burden and additional cost for researchers. While data stewards are expected to be anchored locally in the research environments, it may be advantageous to have national coordination that ensures good practices across the institutions.

Data stewardship competences thus have multiple dimensions, and can be understood as:

1. level-divided academic and professional competence, ranging from qualified librarians or IT staff to specialized BSc, MSc or PhD graduates. The latter may be integrated in the research groups as research competence on an equal footing with researchers' other academic and professional competences (i.e. integrated at peer level as co-author of articles, etc.).

2. having different technical and science content, covering a continuum from general (generic) data management to subject-specific elements.

3. fundamentally different, in both form and content, in the various main areas (technical sciences, natural sciences, health science, social science, and humanities).

Research institutions and DeiC must collaborate on ensuring that data stewardship competences are obtained through upgrading of qualifications and new study programmes.

The research institutions and DeiC must ensure national coordination and facilitate knowledge exchange across research institutions and academic fields.

Research institutions and DeiC must support the use of maturity models and metrics in relation to FAIR, in accordance with emerging international standardization[19].

Proposals for specific actions

● The research institutions support building up local data stewardship support for assistance of the institutions' researchers.

● Coordination and collaboration are established across the research institutions on the provision of data stewardship service where this is beneficial based on, for example, a resource consideration, including for large-scale interdisciplinary projects. Major national actions and initiatives are coordinated under DeiC.

● Relevant continuing education and training in FAIR, data management, and data stewardship must be developed and offered to researchers and data stewards at relevant academic and professional levels.

● Knowledge of data management and the FAIR principles must be made available at relevant stages of the researchers' study programmes, for example in the PhD schools.

---

[19] See, for example, RDA's FAIR Data Maturity Model: specification and guidelines, DOI: 10.15497/RDA0050 and Recommendations on FAIR metrics for EOSC, DOI: 10.2777/70791.

4 Funding

4.1 Principles for distribution of expenses on data management and physical data repositories in connection with externally funded research projects

4.1.1 Research funding

Preparing and maintaining data in accordance with FAIR principles and then making data openly accessible, if there are no explicit arguments to the contrary, requires new research processes, new technical infrastructure and tools, new competences, and support functions. Use of the FAIR data principles is expected to result in better utilization of research investments due to better data management and data re-use. However, the implementation of the strategy does not come for free. The additional costs must be covered by those that already pay for the research, i.e. primarily research institutions and private and public research-funding organizations.

4.1.2 The institutions' obligations

The institutions are obliged to support the researchers with a basic e-Infrastructure, in accordance with what is categorized as *category 2* below, which makes it for example possible to create, use, and store research data securely. The basic infrastructure must be able to handle different types of data and comply with national and international acts and regulations. Competent user guidance and competence building are a fundamental part of the e-Infrastructure.

4.1.3 Funding of and responsibility for establishment of the basic e-Infrastructure

A significant part of the e-Infrastructure will be established through DeiC as a national infrastructure with data warehouses and large-scale computer systems. The whole setup is bound together by the Research Network, which also connects the Danish institutions with international networks through NORDUnet. This provides access to, for example, the European Open Science Cloud (EOSC), where international research institutions and ESFRI projects will make their data accessible in the future. The e-Infrastructure may be funded internationally as part of a research collaboration such as in the ESFRI projects, which are funded through the international memberships.

The costs for the e-Infrastructure offered nationally are paid by the eight universities and the Ministry of Higher Education and Science in accordance with a further agreed distribution scale.

The institutions have full responsibility for the establishment of local IT solutions (see Tables 2a and 2b) to ensure access to national and international resources for researchers and students, as well as to establish support functions where users can get help.

The costs for handling data according to the principles described in this strategy will vary considerably, depending on the size of the projects and the area of focus. Some can naturally be handled by the basic infrastructure, others will, due to scale and complexity, necessitate a significant drain on resources that exceeds what one can expect from the basic e-Infrastructure. It is a prerequisite for the strategy that the

basic infrastructure with adequate supply of resources can be expanded and thus supply the necessary support for large and complex data sets. Only in extremely rare cases should it be necessary to build new IT systems from scratch and to incur expenses for maintenance and operation of these by the institutions.

Basic IT infrastructure for data management will be largely funded through the national collaboration under the auspices of DeiC. In addition, certain research areas already have a significant data infrastructure at international level. Thus further measures will not be need to be taken at national or local level in these areas.

The expenses for large project-specific expansions of the basic infrastructure, in terms of hardware, data management, and increased use of ancillary functions, must be includable in the budgets for the projects. Expenses incurred by the institutions for establishment of the e-Infrastructure may be included as the institutions' co-funding of projects by agreement with the research-funding organization.

4.1.4 Expenses for Data Management and physical data repositories

Expenses for data management infrastructure which the host institutions must establish and maintain to support the projects, can roughly be divided into three principal categories, with some sub-categories:

Category 1—International e-Infrastructure

This category consists of the internationally anchored data warehouses which are funded through international memberships and are maintained by the host institutions, for example EMBL, CERN, ESO, and ESFRI initiated research infrastructures. In many cases, data storage at these institutions will be free of charge, possibly also without Danish membership and payment.

Category 2—Basic e-Infrastructure

Basic IT infrastructure that the research institution must establish, for example networks, print facilities, basic software, security, and the FTEs necessary to keep the basic IT infrastructure functioning and secure. The basic infrastructure also includes data management and general data storage as exemplified in Table 1. The further delimitation of basic IT infrastructure is expected to be made in collaboration between the research institutions and the funding organizations   .

These expenses are categorized as *Category 2.A* and are paid by the institutions, as it must be regarded as a necessity for the maintenance of study programmes and research activities on par with international partners that there is a basic e-Infrastructure. This must be part of the basic budget of the institutions.

However, some projects will draw on the basic IT infrastructures to such an extent that it will be necessary to inject additional resources to cover the e-Infrastructure needs of the project. These may, for example, be general support functions and licences, as well as upward adjustment of storage capacity. Costs related to the increased use in the project of the basic e-Infrastructure and data processing that make data FAIR (but not necessarily open) are characterized as *Category 2.B*, and may be included in the project application budget.

Category 3—Special e-Infrastructure

There are parts of the e-Infrastructure that are solely connected with a given project, for example data stewards attached to the project and the establishment of large new long-term warehouses (repository or databases), and that are not an extension of existing basic e-Infrastructure. These are categorized as *Category 3* and the costs are included in the project application budget.

Table 1 below provides a more detailed description with examples of each element of these categories.

Table 1—FAIR funding categories

Category	Physical infrastructure:	FTEs
1 **International e-Infrastructure** Basic IT infrastructure internationally, funded through international collaboration with the institutions.	Infrastructure provided through international cooperation agreements For example, the European Bioinformatics Institute (EMBL-EBI), CERN, the European ESFRI projects, and other international infrastructures with or without Danish co-funding.	Postings to or secondments with international organizations.

2A **General consumption of basic e-Infrastructure** Basic IT infrastructure nationally or locally, funded by the institutions, possibly through collaboration, for example nationally under the auspices of DeIC.	General IT and storage capacity during and after the end of the grant period (data from the project). The basic e-Infrastructure may be part of national or international collaboration structures, possibly under the auspices of EOSC infrastructure. ESFRI or other international collaboration. Software and services, basic licences, e.g. MATLAB, Microsoft, Amazon, Dropbox, etc.	Operation, maintenance, and development. Basic advice and support. Support functions for researchers, e.g. technical IT, data stewards, GDPR. Cross-institutional harmonization. Monitoring of FAIR principles
2B **Excess consumption of basic e-Infrastructure** Additional IT infrastructure at the institutions connected to research project or research grant.	Very high demand for or draw on data warehouses (repositories or databases). Specific licences. Special software.	Particularly high draws on the institution's IT, research support, or data steward crew, for infrastructure development and data management aimed at making data FAIR.

3	Very special needs for data warehouses (repositories or databases).	Own dedicated Data Stewards for the project and group of researchers.
Special e-Infrastructure	Specific software licences.	
IT linked to a grant or project.	Special software.	

4.2 Principles for establishment of cost level and allocation in one of the two categories.

It must be both transparent and easy for applicants to determine which expenses can be included in the application budget. There may be a difference in what can be applied for from the various research funding organizations, for example due to different practices regarding overheads. The institutions may offer e-Infrastructure on market-like terms.

Proposals for specific actions:

- A list of expenses for FAIR data management-related expenses (software, storage space, FTEs) will be established and updated on an ongoing basis. The list is indicative and can be used by research funding organizations and universities as a starting point for the preparation and assessment of budgets.

- The list will be drawn up by a committee with representatives from the Danish Universities and the research funding organizations.

4.2.1 The research-funding foundations' requirements for handling of research data

Data management requirements are equated with other statutory and contractual rules applicable to externally funded research projects. The host institutions are responsible for compliance with the rules. The researchers thus cannot set aside that it must be assessed to which degree data must be made FAIR even though data are produced as part of an externally funded project. This applies to projects supported by funding from both private and public organizations.

The reporting to the research funding organizations usually comprises a list of publications related to the project. In the future, the list must also include a statement of the datasets established under the project identified by a PID (Persistent Identifier) and with a statement of whether data have been made accessible to other researchers after the results have been published. This is done in accordance with the Danish Open Data Act, as long as it is not contrary to other legislation.

5 Implementation and management

The strategy contains a wide range of recommendations to be implemented. The Danish Ministry of Higher Education and Science will set up an advisory group to ensure progress in the implementation of the strategy. The advisory group must consist of representatives with decision-making competence for the main beneficiaries of the strategy, i.e. research institutions and research-funding organizations. If needed, the advisory group may set up a number of working groups to prepare specific implementation plans. In addition to research institutions and foundations, such working groups may include members from other types of institutions.

The advisory group will report to the Danish Agency for Higher Education and Science and also keep the universities informed about it work.

6 Glossary

Term	Explanation
API	Application Programming Interface (API) is a technical interface directly connected to a computer's software program. It defines the types of calls or requests that can be made, how to make them, the data formats to be used, which data are to be returned, etc.
Data stewardship	Collection, annotation, structuring, meta dating, archiving, and curating of data, with the involvement of policies, legislation, standards, vocabularies, etc., often anchored in a particular academic field
FAIR	**F**indable, **A**ccessible, **I**nteroperable, **R**eusable
FAIR Data Point	Data system that exposes data and metadata in accordance with the FAIR principles
FAIR metrics	Methods for examining the extent to which the FAIR principles have been met. Work is being done internationally, including in EOSC, to establish joint standards for FAIR metrics. FAIR metrics must primarily be seen as a tool available to the researchers to make data more FAIR.
FAIRness	The degree of compliance with the FAIR principles.
Confidential data	Data which cannot be made accessible under an open licence, but which must be kept confidential in respect of intellectual property rights, protection of personal data, privacy, security, and/or legitimate commercial interests.
Metadata	Data describing other data: identification, description/documentation, history of creation, licence, etc.
Maturity models	Maturity models can be used by a research project or subject area to define the desired FAIR objective. FAIR metrics can then be used to evaluate compliance, so that the results are assessed in relation to the targets set.
PID (Persistent Identifier)	Unique identification of a digital resource. Must normally be translated into a specific website address via a PID service (e.g. DOI, ORCID).

Repository   Database of digital objects/research output, comprising data, metadata, and PID, often with a searchable user interface and computer interface (API). Repositories are most often organized by academic fields (disciplinary repository, preferably internationally) or institutionally (institutional repository, typically by a university).

| **Trusted repository** | A repository certified in accordance with a recognized standard, for example Core Trust Seal, which deals with organizational, technical, financial, and legal matters, etc. in relation to the repository. |

7 Appendix 1—Working groups that have contributed

The document has been prepared by working groups set up by DeiC's Board, on assignment from the Danish Agency for Higher Education and Science   .

7.1 FAIR Strategy Steering Committee

- John Renner Hansen, University of Copenhagen, DeiC
- Kirsten Winther Jørgensen, Copenhagen Business School, DeiC
- Anne-Sofie Jensen, Danish National Archives
- Kira Stina Hansen, The Danish Royal Library
- Lars Christensen, Danish Agency for Higher Education and Science
- Anders Sparre Conrad, DeiC

7.2 FAIR Finance Working Group

- John Renner Hansen, University of Copenhagen, DeiC
- Ole Skøtt, University of Southern Denmark, DeiC
- Thomas Midtgaard, Danish Agency for Higher Education and Science
- Johanne Thorup Dalgaard, Independent Research Fund Denmark
- Birgit Pedersen, Independent Research Fund Denmark
- Maja Horst, Independent Research Fund Denmark
- Steen Markus, Danish National Research Foundation
- Søren Degn-Pedersen, Innovation Fund Denmark
- Thomas Brandt Nielsen, Innovation Fund Denmark
- Ulrik Nicolai de Lichtenberg, Novo Nordisk Foundation
- Jan Egebjerg, Lundbeck Foundation
- Lars Hansen, Velux Foundations (Observer)
- Thomas Buchvald Vind, University of Southern Denmark
- Rene Belsø, DeiC

7.3 FAIR Strategy Working Group

- Anders Sparre Conrad, DeiC
- Birte Christensen-Dalsgaard, Aarhus University, DeiC
- Mette Hall-Andersen, Danish National Archives
- Anne Sofie Fink, Danish National Archives
- Bjarne Andersen, The Danish Royal Library
- Thomas Midtgaard, Danish Agency for Higher Education and Science
- Bartlomiej Wilkowsk, Statens Serum Institut
- Ebbe Villadsen, National Research Centre for the Working Environment (NRCWE)
- Martin Sønderholm, Geological Survey of Greenland and Denmark (GEUS)
- Adam Arndt, Danish Agency for Digitisation

- Ivan Thaulow, Statistics Denmark
- Niels Ploug, Statistics Denmark
- Solveig Vibe-Petersen, Statistics Denmark
- Rene Belsø, DeiC

7.4 University Group for the FAIR Strategy

- Birte Christensen-Dalsgaard, Aarhus University
- Susanne den Boer Beckers, University of Copenhagen
- Thomas Schlichting, University of Copenhagen
- Mareike Buss, Copenhagen Business School
- Sacha Zurcher, Roskilde University
- Kirsten Ohm Kyvik, University of Southern Denmark
- Karsten Kryger Hansen, Aalborg University