Preliminary analysis: introduction of FAIR data in Denmark

Produced for the Danish Agency for Science and Higher Education

OXFORD RESEARCH

HØJBJERRE BRAUER SCHULTZ

Preliminary analysis: introduction of FAIR data in Denmark

Purpose of the report

This preliminary analysis examines the potential for implementing FAIR data in Denmark by estimating the benefits and costs of doing so and also by investigating barriers and opportunities in implementing FAIR data in Denmark.

Analysis team

Oxford Research

- Jakob Stoumann, Managing Director
- Morten Larsen, Senior Analyst
- Jakob Falk, Senior Analyst
- Anders Gøgsig Randrup, Analyst

Højbjerre Brauer Schultz

- Esben Anton Schultz, Partner
- Sarah Kildahl Nielsen, Analyst

About Oxford Research

Knowledge for a better society

Oxford Research is a specialised knowledge company focusing on the areas of industrial and regional development and welfare. Oxford Research carries out tailored analyses, implementation evaluations and impact evaluations for public authorities, foundations and organisations in civil society. We also advise on strategy development, facilitate development processes and pass on our knowledge at training courses and seminars. We combine academic depth, excellent communication and strategic understanding – this approach allows us to generate knowledge which is easily applicable and that can make a difference.

About Højbjerre Brauer Schultz

Højbjerre Brauer Schultz is one of the Nordic region's leading socio-economic consultancy firms. Højbjerre Brauer Schultz was founded to provide knowledge that can develop and future-proof the welfare state. We advise public authorities, interest groups, international organisations and private businesses. By building bridges between academic knowledge, empirical results and the political reality, we deliver application-oriented analyses that are well documented and clearly communicated.

Contents

Executive summary

The Danish Agency for Science and Higher Education has commissioned Oxford Research and Højbjerre Brauer Schultz to carry out a preliminary analysis of the potential for implementing FAIR data in Denmark. FAIR is a set of guiding principles to make research data Findable, Accessible, Interoperable and Reusable.

The purpose of the analysis is twofold: 1) to estimate the benefits and costs of implementing FAIR and 2) to examine the barriers and opportunities in implementing FAIR data at the Danish research institutions. Specifically, the analysis provides answers to the following questions:

- What costs and what benefits can be expected from the implementation of FAIR data – is it socio-economically a good case?
- How do the main Danish stakeholders assess the potential of the FAIR data concept? What opportunities and barriers do they see in terms of policy, organisation, technology, new habits/culture, time (for researchers) and service development in implementing FAIR data in Denmark?
- What are the key learnings from the implementation process of FAIR data in Germany in terms of policy, organisation, technology, infrastructure and change in research culture, and what benefits are achieved/foreseen in Germany?

Two main methodological approaches have been used throughout the study: 1) a systematic literature study of the costs and benefits of implementing FAIR data in Denmark and 2) a systematic desk research of relevant studies, reports, articles and 36 in-depth qualitative interviews that focus on key stakeholders' experiences and knowledge of the FAIR principles and their assessment of the potential for introducing FAIR data.

Status – how far has Denmark progressed with FAIR?

Overall, there is no systematic approach to the FAIR principles in Denmark. There is no national Danish policy in this area nor any common consensus on binding guidelines and practices. However, the Danish key stakeholders, including universities, research libraries and political authorities, generally support the FAIR principles and see immense potential for reaping the benefits in terms of better research and socio-economic impact. Currently, the Danish approach to FAIR data is mainly locally based. This means that each university, including its different research departments and faculties, has developed its own approach to data management and thus comply differently with the FAIR principles. Some universities have developed a research data management policy addressing the FAIR principles and invested in data infrastructure that supports a system based on the FAIR principles. Other universities have just recently learned about the FAIR principles and are to varying degrees in the process of developing policies that can support the FAIR principles.

Potential for a positive socio-economic scenario

The literature study indicates that there is a positive socio-economic value in introducing FAIR data in Denmark. This result is based on some of the most conservative assumptions found in the literature. In this scenario we find that the socio-economic net present value is about DKK 2 billion over a 40-year period, if we assume that 50 per cent of all research data in Denmark will comply with the FAIR principles. This corresponds to an annual socio-economic gain of about DKK 50 million.

However, it is important to be aware that the calculations in this study are based on several assumptions and the size of the socio-economic value largely depends on how successful and to what extent the FAIR principles are implemented. Therefore, the socio-economic values should not be regarded as accurate estimates, but rather as an assessment under the chosen assumptions.

Key factors to implementing FAIR data in Denmark

The analysis finds that the implementation of FAIR data – meaning an ecosystem of research data that fosters the findability, accessibility, interoperability and reusability principles – requires that each element in the ecosystem can be linked together and support each other. To have a well-functioning ecosystem to foster the FAIR principles, each element must be handled and balanced. For example, it is a precondition that there should be the right skills to manage and share research data, that there should be a basic infrastructure for storing data, and for researchers to have easy and simple access to data. Since researchers are required to comply with the FAIR principles in practice, it is crucial to stimulate researchers' demand and incentive.

The key factors to implementing FAIR data in Denmark are firstly national coordination and cooperation across research actors, libraries and research funding institutions, because this is necessary for creating coherence in the system. Competence and culture are another key area that must be supported to give the researchers the necessary skills and incentives for sharing data. Access to research data is a third crucial element, because a precondition for encouraging researchers to take part in sharing and using digital research data is to give them easy access to data. In addition, there is a need to address a wide range of technical and legal barriers – not only at national level but also internationally. Finally, the analysis stresses that further investments in specialised infrastructure for storage, handling, processing and dissemination of research data are needed in Denmark.

Recommendation for a Danish FAIR data solution

Overall, we recommend building on existing capacity and expertise. The analysis shows that many of the elements needed to realise FAIR data already exist, but they are fragmented and dispersed. Coordination and collaboration are crucial for developing a common approach and understanding of the FAIR data principles. However, a FAIR data solution should be based on a 'bottom-up' approach, meaning that it should be demand driven and based on the researchers' practices and the international context of research. A FAIR data solution should build on local solutions with a cohesive national superstructure – not a 'one-size-fits-all' solution, but a solution that can grow based on local, academic and research environments.

Summary

Research data has to some extent become part of the agenda. In the first instance, this is due to the ever-increasing volumes of digital research data being generated. Graham Pryor talks about there being a 'flood of data', and has worked out that annual global growth in the production of research data in electronic format is approx. 30 per cent.[1] At the same time, in both policy and research environments, there has been a move towards opening data up; this has resulted in the 'Open Access' and 'Open Science' ambitions.

The European Open Science Cloud (EOSC) was introduced during the Dutch presidency of the EU. The aim of EOSC is to expand digital sharing of research data in Europe. Since then, the FAIR principles (Findable, Accessible, Interoperable, Reusable) have achieved international attention and have gained ground as principles in the Netherlands and Germany. The process of achieving and implementing FAIR in these two countries is in full swing, but experience has shown that it is a complex and comprehensive process which requires focus on many elements, including culture change, infrastructure development and skills development, as well as the reform of scientific practices and merit systems.

The four FAIR principles are as follows:

- **F**indable: Data are assigned globally unique and persistent identifiers so they can be retrieved regardless of the scientific research area.
- **A**ccessible: Data are described based on a standardised communication protocol to make data accessible – even after the data are no longer available.
- **I**nteroperable: Data are interoperable, with a common language used across research fields.
- **R**eusable: Data meet a number of quality standards for legibility – for both people and machines.

The principles are mainly based on the vision of a shared 'data ecosystem', the aim of which is to make digital research data accessible across research actors and academic disciplines. The immediate potential for introducing FAIR appears to be great, as not only will it improve conditions for the individual researcher and increase research quality, but it will probably also create a social benefit, given that the effectiveness of both public and private sector investments in knowledge production is increasing. As our analysis shows, the socio-economic net present value will be approx. DKK 2 billion over a 40-year period, provided that 50 per cent of all research data in Denmark follow the FAIR principles.

Getting FAIR introduced and implemented can be instrumental in giving Denmark a head start internationally as a knowledge and research nation. Denmark is already doing well in an international context in terms of key R&D indicators and is one of the European countries with the highest R&D intensity (measured by R&D share of GDP), but other countries – at the time of writing Germany and the Netherlands – are interested in the opportunities and potential presented by FAIR. This development is something which further emphasises the need for Denmark to take a closer look at the opportunities for integrating the FAIR principles into the management of research data. However, it is also important to weigh up the benefits and costs of implementing such a solution given that, as the analysis shows, it is associated with significant challenges and relatively high costs.

[1] Graham Pryor (2012): Why manage research data?

PURPOSE OF THE ANALYSIS

This preliminary analysis was commissioned in order to examine the potential for implementing FAIR data in Denmark by means of two main elements: 1) estimating the benefits and costs of implementing FAIR research data and 2) identifying and investigating the barriers and opportunities in implementing FAIR research data at research institutions in Denmark. The analysis focuses specifically on answering the questions set out below:

- **What costs and what benefits can be expected from the introduction of FAIR data – is it socio-economically a good case?**
- **What is the process for implementing FAIR research data in Germany, including its organisation, what are the barriers, what investments have been/must be made and what benefits have been achieved/are foreseen?**
- **How does the Danish landscape for production, storage and preservation of research data match up to the FAIR data concept? What opportunities and barriers do they see in terms of policy, organisation, technology, new habits/culture, time (for researchers) and service development? In short, what are the critical factors for implementing FAIR data in Denmark?**

The analysis is based on a number of different methodologies and data sources. This methodological triangulation ensures that the potential for introducing FAIR data in Denmark is illustrated from several relevant perspectives and also guarantees valid and application-oriented analysis results. The estimation of costs and benefits from the introduction of FAIR data in Denmark in this analysis is primarily based on a thorough literature study of benefits and costs associated with the introduction of FAIR data. We have taken a critical view of the literature, and where possible we have used Danish data sources or adjusted the assumptions so they are a better fit for a Danish context. The collection of German experiences and assessment of the Danish opportunities and barriers are based on a combination of a thorough desk research of studies, reports and articles, as well as a wide-ranging interview study focusing on key stakeholders' experiences and knowledge of the FAIR principles and their view of the potential for introducing FAIR research data. We conducted a total of 36 interviews with Danish and German key stakeholders.

STATUS – HOW FAR HAS DENMARK PROGRESSED WITH FAIR?

The Danish key stakeholders, including universities, research libraries and political authorities, generally support the FAIR principles. Several emphasise that it is a matter of sensible principles, something which it is difficult to disagree with. They see great potential for reaping the benefits in terms of better research and socio-economic impact, provided the FAIR principles are adhered to. Most representatives from research institutions do, however, stress that it is difficult to find out exactly what is underneath the individual principles and how they can be put into practice.

Lack of systematic approach and interdisciplinary coordination in the Danish FAIR initiative

Overall, there is not yet a systematic approach to the FAIR principles in place in Denmark. There is no national Danish policy in this area nor any common consensus on binding guidelines and practices for FAIR data. Despite this state of affairs, there is a growing interest in FAIR in Denmark in light of the GO FAIR initiative in the Netherlands and Germany, which has led these two countries to start working actively to promote the agenda to make research data FAIR. Among other things, centralised coordinating Danish actors such as DeIC (Danish e Infrastructure Cooperation), DM Forum (National Forum for Research Data Management) and DEFF (Denmark's Electronic Research Library) have begun taking a closer look at the FAIR principles and what role they can play in a Danish solution for open research data.

Given the lack of a national policy, the Danish FAIR initiative is relatively uncoordinated and many of the actors are unaware of what the general practice is in Denmark in this area. The work on FAIR research data is thus largely locally based. The few existing initiatives that harmonise with FAIR have been instigated by the individual universities (sometimes working with each other) and research libraries. At management level, most institutions have general knowledge of FAIR but no internal, formalised guidelines for sharing and managing data across departments and faculties have been drawn up. This lack of internal formalisation is partly due to great uncertainty about what research data are and how they should be stored. In practice this means that in most places responsibility for sharing and managing data is decentralised to faculty and departmental level, meaning that it is up to the individual researchers to decide on the approach to data management and data sharing. On the other hand, some universities have progressed further in their work with FAIR, including the Technical University of Denmark (DTU), which applies the FAIR principles in its research data management policy for preservation of primary materials and data.

Additional need for investment in the research and data infrastructure

Moreover, there are few Danish universities which make shared systems for data curation available to their researchers. The universities provide storage space for their researchers – often in the form of network drives – but have not made the necessary central investments in either data preservation or integration of data into a system so they can be shared across the university's faculties and departments. There are, however, some individual examples of major system investments. DTU is one institution that has invested in a coherent storage system, currently in the process of being implemented so it can be used for long-term and secure preservation of the university's research data. The University of Southern Denmark (SDU) and Aalborg University (AAU) have also started looking at a shared storage system.

WHAT ARE THE SOCIO-ECONOMIC CONSEQUENCES OF INTRODUCING FAIR DATA IN DENMARK?

An important prerequisite for prioritising the work with FAIR data in Denmark as well as in other countries is FAIR data creating value for society. We have carried out a socio-economic impact assessment that highlights potential benefits and costs of introducing FAIR data in Denmark. In addition to providing an overall estimate of the socio-economic value gained by introducing FAIR data in Denmark, the analysis also provides an understanding of which factors in the organisation and implementation we should bear in mind in order to achieve the best possible socio-economic value from FAIR data.

The starting point for the analysis is the Danish Ministry of Finance's guidelines for producing socio-economic impact assessments[2] and it is based on a thorough literature review of the existing knowledge, as well as analyses of FAIR research data and other open data solutions in other countries.

As a precaution, we decided to base our main scenario on some of the most conservative calculation assumptions found in the literature. We are also disregarding the benefits that would fall to other countries if the FAIR data concept is introduced in Denmark; nor have any benefits gained from a FAIR data collaboration with other countries been included in the calculations. Other expected benefits in the form of new research, greater use of data that would not otherwise be accessible and increased interdisciplinary cooperation within research and innovation have not been included in the calculations either. Consequently, our calculations most likely represent a lower bound estimate of the value of introducing FAIR data in Denmark.

The analysis generally indicates there would be a positive socio-economic value from introducing FAIR data in Denmark. In our main scenario, we find that the socio-economic net present value is approx. DKK 2 billion over a 40-year period, provided that 50 per cent of all research data in Denmark follow the FAIR principles. On average this corresponds to an annual socio-economic benefit of DKK 50 million. If all research data in Denmark follow the FAIR principles, then the present value will be just under DKK 4 billion. The analysis also suggests that the size of the benefit will largely depend on how successfully and to what extent FAIR is implemented. For example, if we include the proportion of research that would or could not be produced without FAIR research data as additional research created as a result of the new data solution, then the socio-economic benefit of introducing FAIR data in Denmark will be significantly greater.

There is a great deal of uncertainty with regard to several of the assumptions on which the calculations are based. Many factors in Denmark may of course be different than what is assumed in the calculations. This means that the calculations must not be considered accurate estimates, rather they must be considered as an assessment of the value of introducing FAIR data in Denmark given the chosen assumptions. If more accurate estimates for the value of introducing FAIR data in Denmark are required, it will be necessary to collect Danish data on a number of the assumptions that form the basis of the calculations.

WHAT ARE THE CRITICAL FACTORS FOR IMPLEMENTING FAIR DATA IN DENMARK?

As we have shown, the analysis indicates that huge economic benefits can be achieved by introducing a FAIR data solution in Denmark. However, there are numerous challenges and critical factors that will need to be addressed for it to be possible to successfully implement FAIR data in Denmark in practice and reap the benefits of increased data sharing and reuse of data that a FAIR solution could potentially bring about.

Generally speaking, the implementation of FAIR research data requires the individual elements of the ecosystem to be coherent and mutually supportive. A properly functioning ecosystem that supports FAIR research data needs these individual elements to be managed and balanced; among other things, it is a prerequisite that the right expertise is in place to manage and share research data, that there is a basic infrastructure for storage and preservation of data and that the researchers must have easy and simple access to

[2] Danish Ministry of Finance (2017)

data. Since it is the researchers who will be implementing and using FAIR data in practice, it is of the utmost importance to stimulate demand from the researchers and use their needs, with regard to reuse and sharing of data, as the starting point. This means that it is not just a question of training and culture but also of providing adequate support for the researchers, including making sure that they have the right search tools, that they can easily access data and that there is a policy setting out their practices.

National coordination between the key actors

Creating coherence in the system means it is it important to have national coordination and cooperation across research actors, libraries and research funding institutions. Interdisciplinary coordination must unite the system and set the overall framework and goals with an eye towards the local level (universities and research libraries), the national level (the complete national infrastructure across universities and disciplines) and the international level (access to the international infrastructure). As described, Denmark does not currently have a national policy or a unifying, coordinating actor to ensure progress in this area; the introduction of requirements for DMPs (Data Management Plans) supported by public sector funding providers may be one way to help drive the introduction of FAIR data.

Researchers must have new skills and incentives to share data

Skills and culture make up another key area that must be supported so as to give researchers the requisite expertise and abilities to manage the sharing of research data. Skills development in Denmark exists locally at some universities but there is no systematic focus on supporting shared skills development, while at several institutions the researchers themselves are responsible for sharing and managing research data without any qualified, local data support. Alongside this, there is a need for a culture change and new understanding among researchers of the fact that research data have research value too; publications are not the only thing that matters. A fundamental challenge here is that the current merit system does not consider researchers who use resources to collect and maintain data, and their time and costs are not acknowledged. Placing more emphasis on the sharing and use of research data in the credit is naturally the most effective way to change researchers' behaviour. Since existing merit practice is international, an international solution is required to change it.

Securing access to research data across research fields

Access to research data is a third, absolutely essential element. A technical solution is required that provides researchers with access to all data across research fields and institutions. The point here is that we are familiar with most of the technical requirements and solutions, and in Denmark there is already an excellent research network in place that connects the Danish universities and research institutions. Yet the main challenge in Denmark is obtaining collected or integrated data and creating connections across departments and faculties at the individual universities where data preservation and data sharing currently take place at departmental level and are not centralised. There is also a need to respond to a wide range of technical, legal barriers – not just at national level but internationally too. Among other aspects, this entails ensuring international standardisation and accessibility and also managing legal grey areas relating to the preservation and sharing of data, e.g. in relation to ownership of data and personal data legislation.

Investments in a specialised infrastructure to support FAIR data

The basis for the ecosystem is a specialised infrastructure for storage, management, processing and dissemination of research data. A well-run infrastructure is a prerequisite for even being able to establish a Danish system for FAIR research data and in Denmark there is a lot to suggest there is still a need for investment to establish data storage and data capacity. Furthermore, in this context it is important to have

long-term and sustainable funding in place for the operation of this storage and the related data services and support functions.

LEARNING AND EXPERIENCES FROM GERMANY

Together with the Netherlands, Germany is one of the countries that has progressed the furthest with FAIR research data. In view of EOSC and the recommendations of the German Council for Scientific Information infrastructures (Rat für Informationsinfrastrukturen, RfII) that has been set up, the GO FAIR initiative is the focal point of the implementation of FAIR data in Germany. The idea behind the GO FAIR initiative is to implement the recommendations of the EOSC High Level Expert Group (appointed by the European Commission) in conjunction with the initiatives and the infrastructure that already exist. The intention of GO FAIR is therefore also for it to be widely supported and consensus-driven based on an acknowledgement that the implementation of FAIR is not just a matter of technical standards but is also largely conditional upon culture, skills development, funding and allocation of responsibility.

Up until now the German initiatives within Open Science have been uncoordinated. They have largely been bottom-up driven but without national coordination. The situation in Germany today is therefore that universities and research libraries in the various federal states have instigated their own initiatives and projects and as a consequence there is not currently a unified overview. The GO FAIR initiative was intended to be a break with existing practice and is a policy commitment to increase interdisciplinary coordination to avoid solutions that are mutually incompatible. The German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) has financed the GO FAIR initiative and there is thus national and central policy support for the promotion of FAIR research data.

In our opinion, Denmark has the opportunity to learn from and build on the German experiences of implementing a FAIR data solution:

Firstly, it is absolutely essential to maintain focus on demand from the researchers as a driver for the initiative. In practical terms, this means that a FAIR data solution must be developed with an appreciation for the researchers' use and needs, e.g. whether they are adequately supported in the changeover to working with FAIR data and whether the solutions developed consider their practices. In Germany they have focused on dialogue with the research environments concerning the potential of FAIR and Open Science; to mention just one example, the GO FAIR initiative focuses on 'maturing' the research environments and engaging in dialogue with the researchers about the potential of FAIR.

Secondly, it is a challenge in Germany to get assured sustainable funding of FAIR. Many of the projects awarded grants have been funded for a limited time period and the risk inherent in project financing is that the projects will founder as soon as the money runs out. In Germany, discussions are ongoing on which funding models people expect to be implemented to ensure long-term and sustainable funding. We believe it to be important for Denmark to also start a discussion on funding models, including how they can contribute to long-term, sustainable funding for infrastructure, training, data support, etc.

Thirdly, experiences in Germany show that the path from making a decision to become FAIR to realising the vision is a long one. The German research system is more fragmented than the Danish system and consists of around 1,000 publicly funded research institutions. At the same time, the German system is federal, meaning that the individual states have policy decision-making authority for research priorities. It must also be expected that it will take a long time to achieve the vision of FAIR data in Denmark, even

though it has a less complex research system and policy decisions are more centralised. This is why in principle a solution can be rolled out more quickly.

RECOMMENDATIONS FOR A DANISH FAIR DATA SOLUTION?

As recommended in the report[3] published by the EOSC High Level Expert Group appointed by the European Commission, it is important to build upon existing capacity and expertise. Many of the elements essential to achieving FAIR data already exist, but the challenge is that they are fragmented and dispersed. The EOSC High Level Expert Group is therefore emphasising the importance of coordination and work across organisations. This recommendation is also supported in this analysis, whose main recommendation is that a Danish FAIR data solution must follow the principle of ensuring balance and creating coherence in the Danish ecosystem so local initiatives are coordinated and do not stand alone.

A demand-driven solution based on what the researchers want

Our assessment, based on interviews with the different Danish stakeholders and the priming experiences from Germany, is that it will be absolutely essential for a Danish FAIR data solution to be demand-driven on the basis of researchers' practice and the international context of the research. In the first instance, the researcher must have support locally at their institution and possess sufficient tools and knowledge to share and manage research data. The framework for the researchers, including the development of policies, incentives, skills and infrastructure, should as far as possible build on international experience and collaboration in the international forums but with the specific needs of the research environments serving as a starting point.

There is a tangible need to look at the researchers' merit system, which does not currently recognise the fact that researchers devote resources and time to collecting and maintaining data. Placing more emphasis on the sharing and use of research data in the credit is naturally the most effective way to change researchers' behaviour. Participation in international forums and partnerships is thus required in order to find a way to manage the basic incentive challenges.

Local initiatives with a cohesive national superstructure

A Danish FAIR data solution should therefore build on the local solutions with a cohesive national superstructure. There is thus a need for national coordination and direction, albeit based on the projects in progress and investments that have already been made. The analysis has shown that some research institutions are already focusing on FAIR and have instigated projects that support FAIR. AAU and SDU, for instance, have come together to invest in a common data storage system that can support the FAIR principles. There is also a certain level of knowledge sharing and inspiration between these two institutions, with SDU looking to DTU for inspiration in its Open Science Policy. It is vital to get these early initiatives into play so that experience and knowledge are not lost and can instead be used in a broader, national solution.

Try to avoid a central 'one-size-fits-all' model – let it be bottom-up

We also believe that it is important for a Danish FAIR data solution not to be a 'one-size-fits-all' solution but one that is allowed to grow based on local, academic and research environments. As is the case in Germany, a national norm should not be dictated for each discipline in Denmark. We should instead fol-

[3] European Commission (2016): Realising the European Open Science Cloud – First report on recommendations of the Commission High Level Expert Group on the European Open Science Cloud

low a bottom-up approach, with the institutions embracing the FAIR principles so we can use this as a basis for discussing how the principles can be translated into practice. The standards developed for sharing research data must meet researchers' needs, including the different nature and specifications of the various research disciplines. It is therefore essential for the researchers, with sufficient institutional support from their university and research libraries as well as through shared, national coordination, to serve as the drivers to mature and refine standards and solutions.

1. Introduction

In recent years, both internationally and within Denmark, there has been greater focus on the potential for more openness in the scientific process (Open Science). Open Science can be seen as a migration from dissemination of research solely through articles for scientific journals towards sharing and use of all accessible knowledge earlier on in the research process. Open Science leads to knowledge being made directly accessible and open to a wider public, including researchers, citizens, institutions and companies, to a much greater degree. This applies both to scientific publications (Open Access) that are typically the final results of the research process and to data (Open Data) worked on by the research. The greater accessibility of publications and data makes the research process far more transparent. This has a positive effect on the integrity of the research as there are better opportunities to reproduce and test the results of research. There is special focus on three elements of Open Science within Denmark:

- Open Access for scientific publications
- Open Data (open research data)
- Research integrity.

This analysis relates to open research data. The increased level of research activity and the digitalisation of research knowledge and data mean that huge volumes of data are now being generated within all fields of research. Consequently there is a need for an initiative here, partly to utilise the huge potential in data and partly to streamline and improve the quality of research by increasing the opportunities for cooperation and reducing the costs of data collection as a result of the use of data that would not otherwise be accessible for reuse.

As part of the European Commission's plans to set up a European Open Science Cloud (EOSC), where the idea is to create open access for 1.7 million European researchers to research data in Europe, FAIR (Findable, Accessible, Interoperable, Reusable) is the 'concretisation' of principles that may form the groundwork for a new way of using digital research data. The FAIR principles do not specify technical requirements; they are in actual fact guidelines for greater reuse of digital research data. In addition to this, an absolutely key point here is that the principles do not support a single standard but are instead indicative and pave the way for different operational approaches.

The purpose of this analysis is to investigate the potential for implementing a FAIR research data solution in Denmark. There is great immediate potential for introducing FAIR, as a full implementation of FAIR research data would not only improve conditions for the individual researcher but would to a great extent also create a socio-economic benefit, given that the effectiveness of both public and private sector investments in knowledge production is on the rise. On the other hand, a FAIR data solution in Denmark has to be implemented in a complex context with different stakeholders and needs. Such a solution will also require massive investments in infrastructure, competence building and data supporters.

To ensure a knowledge-based foundation for this action, the Danish Agency for Science and Higher Education has therefore asked Oxford Research and Højbjerre Brauer Schultz to carry out a preliminary analysis of the introduction of FAIR research data in Denmark focusing on: 1) estimating the benefits and costs of implementing FAIR and 2) investigating the barriers and opportunities in implementing FAIR research data at research institutions in Denmark. Specifically, the analysis will answer the following questions:

- **What costs and what benefits can be expected from the introduction of FAIR data** – is it socio-economically a good case?
- **What is the process for implementing FAIR research data in Germany**, including its organisation, what are the barriers, what investments have been/must be made and what benefits have been achieved/are foreseen?
- **How does the Danish landscape for production, storage and preservation of research data match up to the FAIR data concept?** What opportunities and barriers do they see in terms of policy, organisation, technology, new habits/culture, time (for researchers) and service development – in order to implement FAIR data in Denmark?

1.1 HOW TO READ THIS REPORT

As an opening to the two interdisciplinary analyses, the report will set out below a brief overview of the background to FAIR, as well as whether and to what extent Denmark is already working with FAIR data. Chapter 2 presents the first interdisciplinary analysis – an analysis of the costs and benefits of introducing FAIR data in Denmark. Chapter 3 presents the second interdisciplinary analysis – an analysis of the barriers and opportunities in introducing FAIR data in Denmark. Chapter 4 provides an overview of the experiences with FAIR research data in Germany. The methodological considerations and underlying data of the analysis are described in Chapter 5, while Chapter 6 lists the report's references.

1.2 BACKGROUND TO FAIR

In general terms, the FAIR principles outline the conditions for good data management. The principles constitute an attempt to define a set of generally valid criteria for appropriate data sharing. The criteria must ensure transparency through clearly specified requirements for (meta)data that constitute a prerequisite for the practical reusability[4]. Data must here be taken to mean digital information that is used to produce knowledge in different fields of research and can thus assume countless forms. The only practical definition is the common digital format that constitutes a prerequisite for sharing on online platforms[5].

FAIR covers four principles which in Danish are referred to as 'genfindelig' (findable), 'tilgængelig' (accessible), 'kompatibel' (interoperable) and 'genanvendelig' (reusable). The conceptualisation and operationalisation of the principles are described in the table below.

[4] Wilkinson et al. (2016): The Fair Guiding Principles for Scientific Data Management and Stewardship.
[5] Force11 (2017): Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing

Table 1.1 Overview of the FAIR principles

FAIR principles	Conceptualisation	Operationalisation (requirements)
Findable	It must be possible for data to be found by machines since the volumes of data are too large for humans to handle.	• Data are assigned globally unique and persistent identifiers (O-ID). • Data are described with exhaustive metadata and are registered or indexed in a searchable database.
Accessible	Data must be accessible – but not necessarily by everyone.	• Data can be retrieved on the basis of O-IDs, which are described according to a standardised communication protocol. • Metadata are accessible – even after the data are no longer available.
Interoperable	Data must be interoperable, i.e. combinable with other data sets.	• Data make use of a formal, accessible, widespread and widely used language. • Data include relevant references to other data sets.
Reusable	It must be possible to reuse data in future research projects and then process these data further.	• Data describe multiple precise and appropriate properties. • Data are released with a clear and accessible data licence. • Data are connected to their origin. • Data meet standards relevant to the field.

The FAIR principles stem from a voluntary stakeholder network, which in very little time has managed to shape the Open Science policy agenda. The principles have rapidly attracted international attention and gained ground as established standards at several EU institutions. Among other successes, the FAIR principles have become an integral part of the European Commission's work on a European Open Science Cloud (EOSC). EOSC is an initiative that follows up on the European Commission's strategy for the EU's digital internal market, and aims to achieve optimal reuse of research data. EOSC will act as a liaison between national initiatives and establish the framework for the creation of a common European infrastructure. It is thus a matter of a European superstructure on top of existing national systems in the form of uniform standards for professional data management, expert advice, funding, best practice examples and technical support.

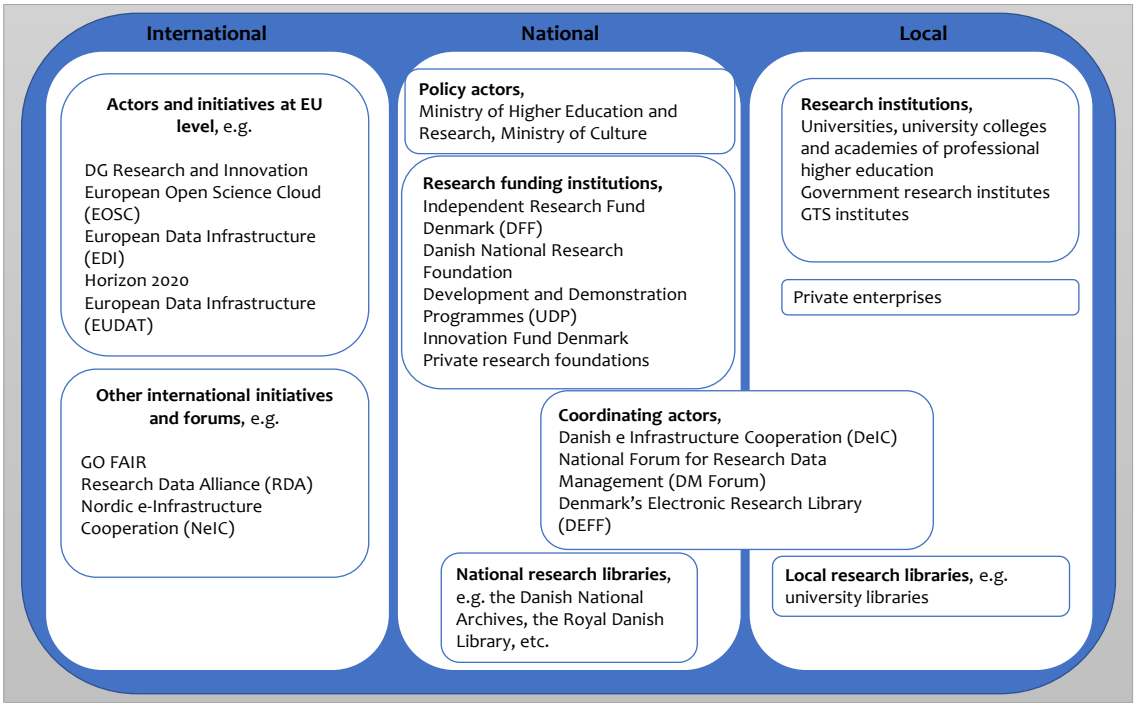
The G20 has also signed up to the principles, while both Science Europe and the EU's research and innovation programme Horizon 2020 have grasped the potential of more efficient use of research data; this is why they are now imposing explicit requirements for grant recipients to follow the logic of the FAIR principles.

1.3 HOW FAR HAS DENMARK PROGRESSED WITH FAIR?

The analysis shows that Denmark is not currently working systematically with the FAIR principles. This was also to be expected since Denmark has neither a national policy in this area nor any common binding guidelines and practices for FAIR data. However there are a series of initiatives – nationally and locally – that may help to underpin the FAIR principles.

The Danish and international stakeholder landscape for FAIR research data consists of a number of stakeholders with different roles and interests. As shown in Figure 1.2 below, there are local, national and international levels when it comes to FAIR research data.

Figure 1.2 Key stakeholders in relation to FAIR data



The FAIR principles originate from an international agenda and as described have become an integral part of the European Open Science Cloud (EOSC). DG Research and Innovation under the European Commission has funded the EOSC pilot project, which is intended to support the migration to EOSC. Using the EOSC High Level Expert Group's recommendations in its report 'Realising the European Open Science Cloud' as its starting point, the GO FAIR initiative was launched with the Netherlands and Germany as the leading participants. Its purpose is to clarify the EOSC policy within three pillars, which are GO CHANGE (culture), GO TRAIN (skills development) and GO BUILD (technical implementation) respectively. There are also a range of international networks and initiatives in place of relevance to FAIR and in which Denmark is an active participant, including EUDAT (European Data Infrastructure) and RDA (Research Data Alliance).

At a national level there is a wide spectrum of different stakeholders in the area. At policy level both the Danish Ministry of Higher Education and Science and Ministry of Culture Denmark are major actors when it comes to ensuring that the research e-infrastructure is coordinated. The research funding institutions, including Independent Research Fund Denmark, Danish National Research Foundation and Innovation Fund Denmark, are also major players in this area given that they fund a major part of Danish research and thus research data too. At the same time, a number of national research libraries, including the Royal Danish Library, the Danish National Archives, etc., by virtue of their role in collecting, processing, storing and making research data available play an absolutely key role in this area. The Danish Ministry of Higher Education and Science, the Danish Ministry of Education and Ministry of Culture Denmark have jointly funded Denmark's Electronic Research Library (DEFF) as a way of promoting the coordination of electronic research libraries. DeIC (Danish e-Infrastructure Cooperation) has similarly been founded to support Denmark as an e-Science nation through delivery of e-infrastructure (computing, data storage and networks) for research and thereby constitutes an absolutely key actor in FAIR data.

The local level hosts the implementing actors, i.e. researchers who actually carry out research or use research results for their own purposes. It is the researchers who ultimately operationalise and use the FAIR principles through their practice. Enterprises also have a significant interest in obtaining access to research, given that it contributes to their development and growth. The researchers carry out research at Danish universities and at the GTS and government research institutes. In relation to the Danish research institutions, the research libraries serve an important role by supporting the documentation and storage of research data.

1.3.1 Danish focus on Open Access and Open Science

Since 2012 Denmark has actively participated in the trend towards Open Science[6] within three research areas [7]:

- Open Access
- Open Data
- Research integrity.

In 2012, the public sector research councils and foundations (Danish National Research Foundation, Independent Research Fund Denmark and Innovation Fund Denmark) published a common Open Access policy that resulted in researchers and other recipients of grants from research councils having to publish their research results with Open Access. The policy is a so-called green Open Access policy, which means that the researchers must parallel archive – if the journals allow this – their research articles in an institutional or subject-specific repository, i.e. a digital archive, to which there will be open access online for all interested parties.

[6] "There is no formal definition of open science. In this report, the term refers to efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation." (OECD 2015).

[7] Danish Government (2016), Science, Technology and Innovation Policy report

In 2014 this was followed by a National Strategy for Open Access, which underpins the 'green' path to Open Access. A National Steering Group for Open Access has been set up to coordinate and support implementation of the national strategy with representatives from all the Danish universities, the Assembly of Director Generals of the Danish Government Research Institutes, University Colleges Denmark, Independent Research Fund Denmark, the Lundbeck Foundation, the Novo Nordisk Foundation and Denmark's Electronic Research Library. Among other resolutions, the Steering Group decided to establish an Open Access Indicator to regularly monitor the progress of Open Access in Denmark. For 2017 the indicator shows that 23 per cent[8] of the scientific articles from research institutions were Open Access.

Open Data means that publicly funded research data are made accessible in digital format with no or minimal restriction. One aspect of Open Data is that the concept is a central part of the European Commission's plans to establish the European Open Science Cloud (EOSC); one of its aims is to develop an open community for all 1.7 million European researchers where they can store, share and reuse research data. The FAIR principles are thus closely linked to the Open Data intention, but with exceptions based on the principle: *As Open as Possible, as Closed as Necessary*. As previously mentioned, Denmark does not have a national policy in this area, although several local initiatives have been launched by individual universities and research libraries that support the trend towards open data.

Part of the logic behind Open Science is that the greater accessibility of publications and data makes the research process more transparent, something which has a positive effect on the integrity of the research as there are better opportunities to reproduce and test the results of research. In conjunction with this and together with Universities Denmark, the Danish Ministry of Higher Education and Science has launched the Danish Code of Conduct for Research Integrity. The Code contains basic principles for integrity (honesty, transparency and accountability) and standards for responsible research practice covering areas such as data management. It says in the Code that *"data should be retained, stored and managed in a clear and accurate form that allows the result to be assessed, the procedures to be retraced and – when relevant and applicable – the research to be reproduced."*[9] All Danish universities plus many different research councils and private research foundations, as well as GTS and government research institutes, have signed up to the Code.

1.3.2 Growing Danish interest in FAIR data

Our interviews with heads of department at Danish research institutions show that data management policies are becoming part of the agenda, with most research institutions already having defined guidelines and practices for how researchers handle data. The interviews demonstrate that in individual cases the FAIR principles have been specified in the policy, but it is still unclear how researchers' practice is going to comply with the principles given that they often lack access to shared data storage, do not have the resources or do not receive adequate data support to make the principles a reality.

At national level, the interviews with DeIC and DEFF show that people here have started to look more closely at FAIR and what role they can play in the Danish FAIR data solution. For example, in its work to promote the development of a network of electronic libraries, DEFF incorporated FAIR into its evaluation criteria for the awarding of small and large project grants to Danish research libraries. In 2014, DEFF and DeIC also formulated a national strategy for research data management, in which its purpose is stated to be: *"to ensure Denmark better and more competitive research through efficient collection, securing, dissemination and reuse*

[8] The Danish research database, http://www.forskningsdatabasen.dk/
[9] Danish Code of Conduct for Research Integrity, p. 9

of relevant research data[10]." Since 2015 the strategy has supported a joint effort on data management, including collaboration in a national forum for research data management (DM Forum) and local initiatives such as the establishment or expansion of support functions at universities and other research institutions. DM Forum is currently trying to incorporate elements of the international development that has taken place since the strategy was formulated; at the European level this means the FAIR principles and EOSC in particular. According to the Chairman of DM Forum, Anders Sparre Conrad, the aim here is to create the necessary coherence between local, national and international initiatives.

A review of the interviews with heads of department from the research institutions generally shows that there is a growing interest in FAIR, seen in the light of more research being data-driven, and that research funding institutions, particularly for Horizon 2020, are imposing data management requirements. Yet not all heads of department know about the FAIR principles. Some of them were not aware of FAIR until we contacted them as part of our work on the analysis. Others are in the process of integrating the FAIR principles in their policies, while some have already incorporated the FAIR principles in their policies. Open Access and Open Science have been on the agenda at many universities for quite some time, which is one of the reasons why the Technical University of Denmark (DTU) and the University of Copenhagen (UCPH) have invested in shared systems and infrastructure that can support the implementation of FAIR.

Table 1.2 below gives an overview of the Danish research institutions' focus on FAIR and examples of FAIR-related initiatives that we have identified. The table has been prepared on the basis of 30 interviews with deans, heads of department, employees and researchers at the Danish research institutions and other key actors in this area. The table is not a systematic review of all initiatives, rather an overview of the policies and initiatives cited as examples by the interviewees.

Table 1.2. Danish research institutions' knowledge of FAIR and examples of initiatives that support FAIR

Institution	Knowledge of FAIR	Examples of FAIR-related initiatives
University of Copenhagen (UCPH)	Knowledge at management level, but most of the researchers do not know about the FAIR principles. One of the effects of an application requirement for data management plans in H2020 is that awareness of FAIR is increasing.	• ERDA – KU Science has implemented a repository in which researchers can store data and make them accessible to other KU Science researchers. • Dataverse – a collaboration with the Royal Danish Library and KU HUM (humanities at UCPH) on the dissemination and accessibility of research data.
Aalborg University (AAU)	A relatively strong focus on the FAIR principles at management level – FAIR to be integrated as part of the research.	• Working with SDU to build up data infrastructure/storage capacity that can support the FAIR principles.
Aarhus University (AU)	Did not know in advance about the FAIR principles specifically, but there is an increasing focus on Open Science/Open Data, including as a result of requirements for data accessibility from multiple journals.	• 'Open Science platform' – AU has established a platform where industry and research can share data and knowledge.
Technical University of	FAIR is a priority area at management level and knowledge of the FAIR principles is growing	• 'DTU Research Data Management Forum' – since 2015 DTU has established a central

[10] Danish National Strategy for Research Data Management 2015–2018, p. 3

Denmark (DTU)	among researchers.	office for tasks including making data FAIR. • Data repository – for some years DTU has invested in an interconnected repository that is expected to be implemented within the next two to three years.
University of Southern Denmark (SDU)	Open Access has been a major focus area for a long time. There is increasing focus on expanding this to Open Science, which incorporates the FAIR principles.	• 'SDU Open Science Policy' – has developed an open science policy containing the FAIR principles. • Working with AAU to build up data infrastructure/storage capacity that can support the FAIR principles.
Copenhagen Business School (CBS)	Knowledge of FAIR at management level, but the researchers generally do not know about FAIR. Some researchers do have experience of using external (foreign) repositories for sharing data.	• 'CBS Research Data Management Policy' – refers to the FAIR principles in the data management policy.
Roskilde University (RUC)	Knowledge of FAIR at management level, but generally limited knowledge of FAIR among researchers.	• Has developed a policy for research data management, in which parts of the FAIR principles are included implicitly.
IT University of Copenhagen (ITU)	Knowledge of FAIR. ITU is in the process of developing a research data management policy which, among other things, contains the FAIR principles. This is expected to be ready by the end of 2018.	
The Royal Danish Academy of Fine Arts Schools of Architecture, Design and Conservation (KADK)	Did not know in advance about the FAIR principles, but experiencing increasing interest in sharing data.	
Danish National Archives	Aware of FAIR – feels that the principles are generally a large area of focus, as FAIR is 'classic archival thinking'.	• Currently preparing a digitisation strategy with special focus on research data.
Royal Danish Library	Aware of FAIR, but not working actively to integrate the principles.	• 'Data management in practice' – has just completed a pilot project in which a data repository was established and made available to all Danish researchers.

This review shows that there is a relatively strong focus on the open research agenda and in some cases specifically on FAIR. The various initiatives, in which some people work explicitly with FAIR while others work implicitly with initiatives that support the sharing of data, show that the efforts are relatively uncoordinated and largely locally based. This means that the management and sharing of data vary widely across departments and faculties.

2. The socio-economic consequences of introducing FAIR research data in Denmark

An important prerequisite for prioritising the work with FAIR research data in Denmark as well as in other countries is FAIR data creating value for society. In this chapter we have carried out a socio-economic impact assessment that highlights potential benefits and costs of introducing FAIR data in Denmark. In addition to providing an overall estimate of the socio-economic value gained by introducing FAIR data in Denmark, the analysis also provides an understanding of which factors in the organisation and implementation we should bear in mind in order to achieve the best possible socio-economic value from FAIR data.

The starting point for the analysis is the Danish Ministry of Finance's guidelines for producing socio-economic impact assessments[11] and it is based on a thorough literature review of the existing knowledge, as well as analyses of FAIR data and other open data solutions in other countries.

As a precaution, we decided to ground our main scenario on some of the most conservative calculation assumptions found in the literature. We are also disregarding the benefits that would fall to other countries if the FAIR data concept is introduced in Denmark; nor have any benefits gained from a FAIR data collaboration with other countries been included in the calculations. Other expected benefits in the form of new research, greater use of data that would not otherwise be accessible and increased interdisciplinary cooperation within research and innovation have not been included in the calculations either. Consequently, our calculations most likely represent a lower bound estimate of the value of introducing FAIR data in Denmark. For example, if we include the proportion of research that would or could not be produced without FAIR research data as additional research created as a result of the new data solution, then the socio-economic benefit of introducing FAIR data in Denmark will be significantly greater.

The analysis generally indicates there would be a positive socio-economic value from introducing FAIR data in Denmark. In our main scenario, we find that the socio-economic net present value is approx. DKK 2 billion over a 40-year period, provided that 50 per cent of all research data in Denmark follow the FAIR principles. On average this corresponds to an annual socio-economic benefit of DKK 50 million. If all research data in Denmark follow the FAIR principles, then the present value will be just under DKK 4 billion. The analysis also suggests that the size of the benefit will largely depend on how successfully and to what extent FAIR is implemented.

There is a great deal of uncertainty with regard to several of the assumptions on which the calculations are based. Many factors in Denmark may of course be different than what is assumed in the calculations. This means that the calculations must not be considered accurate estimates, rather they must be considered as an assessment of the value of introducing FAIR data in Denmark given the chosen assumptions. If more accurate estimates for the value of introducing FAIR data in Denmark are required, it will be necessary to collect Danish data on a number of the assumptions that form the basis of the calculations. Chapter 5 contains a more detailed description of the methodology and data behind the calculations.

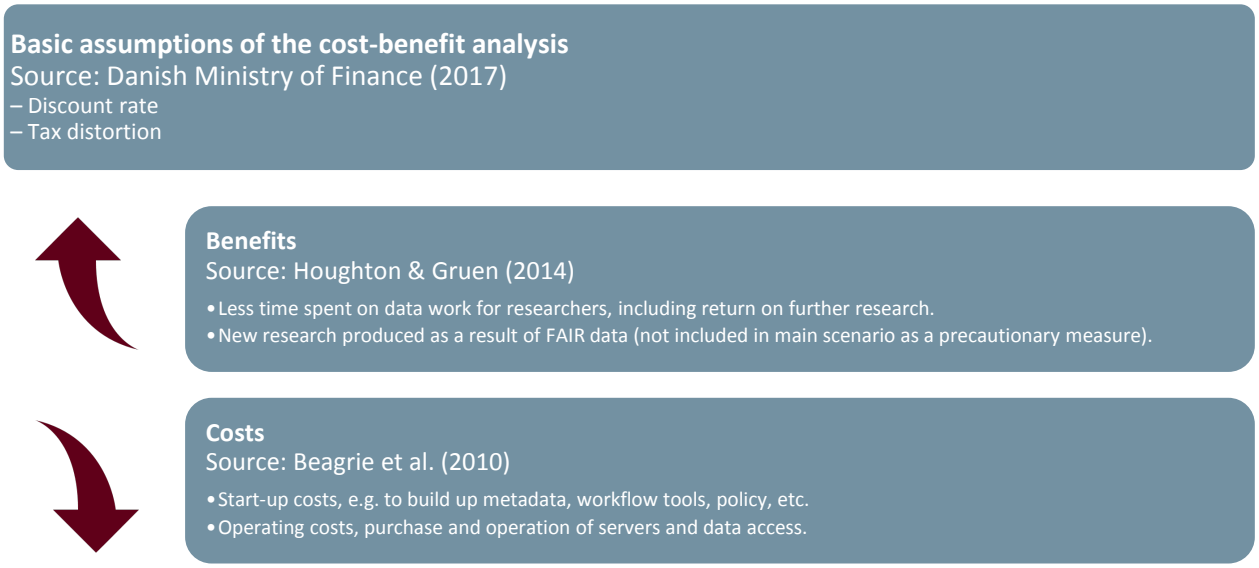
[11] Danish Ministry of Finance (2017)

In the following section, we will present the overall model on which our analysis is based. Then we will present results from our main scenario before finally presenting a number of illustrative calculations of the consequences of changing some of the most critical assumptions in the analysis.

2.1 ANALYSIS MODEL

The analysis is a socio-economic impact assessment (also called a cost-benefit analysis) that follows the Danish Ministry of Finance's methodical guidelines. The starting point for the benefits set out in the analysis is the Houghton & Gruen (2014) model, while the costs from introducing FAIR data are calculated based on Beagrie et al. (2010), cf. the figure below.

Figure 2.1. Analysis model



The main scenario in the analysis has an analysis period of 40 years. For the first 20 years, FAIR data will be implemented gradually in Denmark so that for the next 20 years there are fully implemented FAIR data. According to the Ministry of Finance's guidelines, the analysis period ought to correspond to the expected useful life of the investment. In our analysis, the investment consists of the costs for developing data solutions, guidelines, metadata, etc. There is great uncertainty associated with the useful life of the investment. Once the solution has been developed, the expectation is that it could exist for at least 20 years with minor updates and adjustments as part of normal operation. One barrier that will affect the analysis period is the researchers' mindset and the culture of protecting research data that will be exceptionally difficult to change. Section 2.3 contains a sensitivity analysis of the assumption about the analysis period. Barriers related to culture change can also have some bearing on the proportion of data that can be converted to FAIR research data; this is why in the main scenario we have calculated both the value when 50 per cent of all research data in Denmark are FAIR and the value when 100 per cent of all research data are FAIR.

In the socio-economic impact assessment, we want to evaluate the overall socio-economic value for Denmark throughout the analysis period. It is therefore important to be able to compare benefits and costs that are realised in different years during the analysis period. There are many theoretical and empirical arguments for benefits and costs that will be realised in the future having to be given a lower weighting than benefits and costs realised today. That is why we cannot simply add together the immediate values of benefits and costs over time. The Ministry of Finance recommends that future values be scaled so that the typical preference for current consumption is reflected rather than future consumption. This is done by discounting future values using the so-called discount rate.

We've applied a number of key parameters as recommended in the Ministry of Finance's guidelines, including the discount rate and tax distortion factor. Our results have been converted into a set of indicators, also as recommended by the Ministry of Finance. figure 1.2. contains a more detailed description of the main parameters and indicators used in our calculations.

Figure 1.2. Main parameters and indicators in socio-economic impact assessments

Discount rate: The socio-economic discount rate is used to weight the future benefits and disadvantages in such a way that they can be compared directly to present values. The rate is composed of the risk-free real rate of interest and a supplement for systematic, non-diversifiable risk. The discount rate follows a falling profile over very long periods of time.

Tax distortion factor: The tax distortion factor is used when calculating the socio-economic consequences of initiatives where the public sector bears the costs (or accrues the income). The factor is based on an assumption that public consequences will lead to the tax going up or down. The tax distortion factor represents the marginal social cost which arises because of the need for general tax financing.

Net present value: The socio-economic net present value is the sum of the discounted benefits and costs. The net present value thus expresses the total value of the project to society. A project is profitable if the net present value is positive.

Benefit/cost ratio: The total discounted benefits divided by the total discounted costs. This figure expresses the robustness of either a positive or negative net present value.

Return ratio: The net present value divided by the discounted net investment. This figure expresses the total socio-economic return on the total investment.

Internal rate of return: The internal rate of return indicates the annual socio-economic return on the investment and represents the interest rate at the point where the net present value becomes zero.

2.1.1 Socio-economic benefits

The socio-economic benefits in our analysis include the time saved by researchers. We have also performed sensitivity analyses on the importance of new research that is produced when data are FAIR.

Experiences from data centres with FAIR data in other countries indicate that making data FAIR can lead to significant time savings for researchers who work with data. These time savings are produced thanks to access to data becoming easier and/or the researchers not having to generate data themselves. Our analysis thus includes the value of the time saved by the researchers. This is calculated based on international studies of how much time the researchers can save when research data are FAIR and given a value based on researchers' salaries. In addition, we assume that the time the researchers save will be spent on doing more research, which will lead to an additional socio-economic return. The return is calculated based on studies of the return on research and the previously calculated direct value of the time saved. The value of the time saved is estimated at DKK 10.5 billion over a 40-year period in our main scenario (for more details on the calculation method, see Chapter 5).

Experiences from data centres also indicate that some centre users would not have been able to complete their research if data centres did not exist. In other words, FAIR data are helping to produce new research. This may be related to the fact that researchers do not become aware of data until they are made FAIR, that research projects with smaller budgets can now afford to use data too or that interdisciplinary collaborations will be easier with FAIR data. We have disregarded the value of new research in our main scenario. The reason for this approach is that the existing literature does not contain solid knowledge of the extent to which the new research displaces other research or gives rise to less research in other fields. It may be that smaller amounts of data will be collected since there is now a greater supply of existing data, or simply that the researchers' limited time will move from one type of research to another. A conservative assumption is therefore that new research merely displaces existing research. An equivalent assumption is that the value of the new research is equal to the value of the research that is displaced or disappears when data are made FAIR. However, we can expect there to be a tendency towards researchers wanting to spend their time on the research that gives the biggest return. Completely disregarding the value of new research means that we have a conservative estimate of the socio-economic value of FAIR data.

Due to too great uncertainty or insufficient data, the analysis also disregards several other potential benefits of introducing FAIR data in Denmark, see Chapter 5.

2.1.2 Socio-economic costs

In our analysis, the costs comprise start-up costs and operating costs. The existing studies of FAIR data or other open data solutions have typically focused less on cost than on benefits. The costs set out in our analysis are therefore based on a single study[12] and are subject to great uncertainty. Yet in our opinion the costs express an upper bound estimate for the potential costs, cf. Chapter 5.

Start-up costs cover analyst work in connection with the start-up, building metadata and frameworks for metadata, developing tools and software and also preparing guidelines and any legislation in connection with the introduction of FAIR data. On average they are estimated to account for 24 per cent of the total

[12] Beagrie et al. (2010)

wage costs for data during the implementation period and are expected to rise in the period. The assumption is that start-up costs will rise during the period, since we can expect the data which are easiest to make FAIR will be made FAIR first. Taken as a whole, start-up costs throughout the implementation period represent a net present value of DKK 6.1 billion in the main scenario.

Operating costs cover local data management costs such as the purchase and operation of servers, as well as online access, backup and long-term preservation of data. They are estimated to account for 3 per cent of the total wage costs for data for the whole period, although they are expected to rise in the implementation period in proportion to the amount of data made FAIR. Operating costs during the 40-year period represent a net present value of DKK 0.8 billion in our main scenario.

2.2 RESULTS IN MAIN SCENARIO

As described above, our general assessment is that the results in the main scenario represent a conservative estimate of the value of introducing FAIR data in Denmark. This is due to the fact we have disregarded a number of possible benefits, potentially overestimated the costs and have also chosen to base the calculations on the most conservative assumptions that we identified in the literature. It is worth noting that some of the calculation assumptions applied may be optimistic in a Danish context but these were only used in cases where we were unable to find other estimates in the literature. Section 2.3 contains several sensitivity calculations that illustrate the consequences of changing these calculation assumptions.

Our calculations generally indicate there would be a positive socio-economic value from introducing FAIR data in Denmark. This presupposes that FAIR data will significantly reduce the amount of time that researchers devote to data work. If this does not happen, the introduction of FAIR data could lead to a socio-economic loss. The socio-economic value of FAIR data is thus very sensitive to this parameter, making it important to maintain focus on achieving the greatest efficiency gains possible should it be decided to introduce FAIR data in Denmark.

In our main scenario, we find the introduction of FAIR data to have a socio-economic net present value of between DKK 2 billion and DKK 4 billion over a 40-year period. This equates to an average annual gain of DKK 50–100 million. The value depends on the proportion of all data made FAIR. Were, for instance, 50 per cent of all research data in Denmark to be made FAIR, then the net present value would be DKK 2 billion, compared to just under DKK 4 billion if 100 per cent of all research data were made FAIR. This means that the socio-economic benefit increases as more data are made FAIR. What proportion of the Danish research data that can be made FAIR will depend on a whole host of factors, the descriptions of which are one aspect of Chapter 3. Several FAIR data initiatives are already in place in Denmark and it is not particularly realistic to make all research data FAIR. This means it is not possible to achieve benefits equivalent to 100 per cent of all data being made FAIR.

The result in our main scenario is relatively robust (regardless of whether it is 50 per cent or 100 per cent of all data that can be made FAIR) because the benefits are approx. 40 per cent greater than the costs. The internal rate of return indicates that the annual return on the investment in FAIR data is around 7 per cent, which is significantly higher than the discount rate of 4 per cent. The return ratio shows that the investment in FAIR data (measured here as the start-up costs) generates a socio-economic value around half the size of the investment, which is relatively low compared to estimates for return on investment (RoI) for investments in data, library and information services. The estimates for these types of investments in-

dicate that they create value between approx. three and five times the size of the investment.[13] The very high costs in our analysis are the most likely cause of this. The remaining indicators are not dependent on what proportion of all data that can be made FAIR. This is because both benefits and costs are expected to grow exponentially, meaning that not only the largest benefits by far but also the largest costs by far are realised at the end of the analysis period.[14] In the following figures it is assumed that 75 per cent of all data will be made FAIR.

Figure 2.2. Socio-economic indicators – main scenario

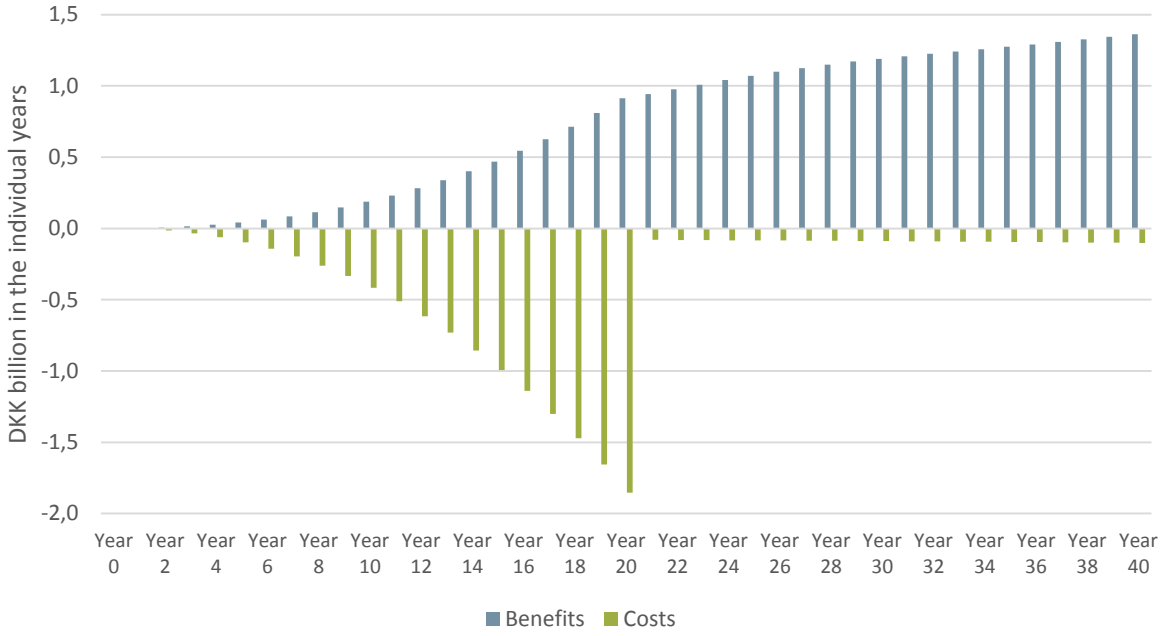
Net present value: DKK 2.0–3.9 billion	Benefit/cost ratio: 1.4	Return ratio: 0.5	Internal rate of return: 7 per cent

Note: The low net present value is if 50 per cent of all research data in Denmark are made FAIR, while the high value is for 100 per cent. The return ratio is calculated as the net present value divided by the discounted value of the start-up costs. The figures are rounded.
Source: Own calculations.

Given an implementation period of 20 years and a useful life of 20 years for the FAIR data solution, there will be greater costs than benefits from FAIR data in the first 20 years. From then on, however, the benefits significantly exceed the costs, cf. figure 2.3.

[13] Beagrie (2012)
[14] This assumption is based on the literature emphasising that the benefits of open data increase exponentially due to the network or system effects associated with data – data will be worth more if more data are accessible. On the other hand, with regard to the costs it can be assumed that when implementing FAIR data the data easiest to make FAIR will be dealt with first.

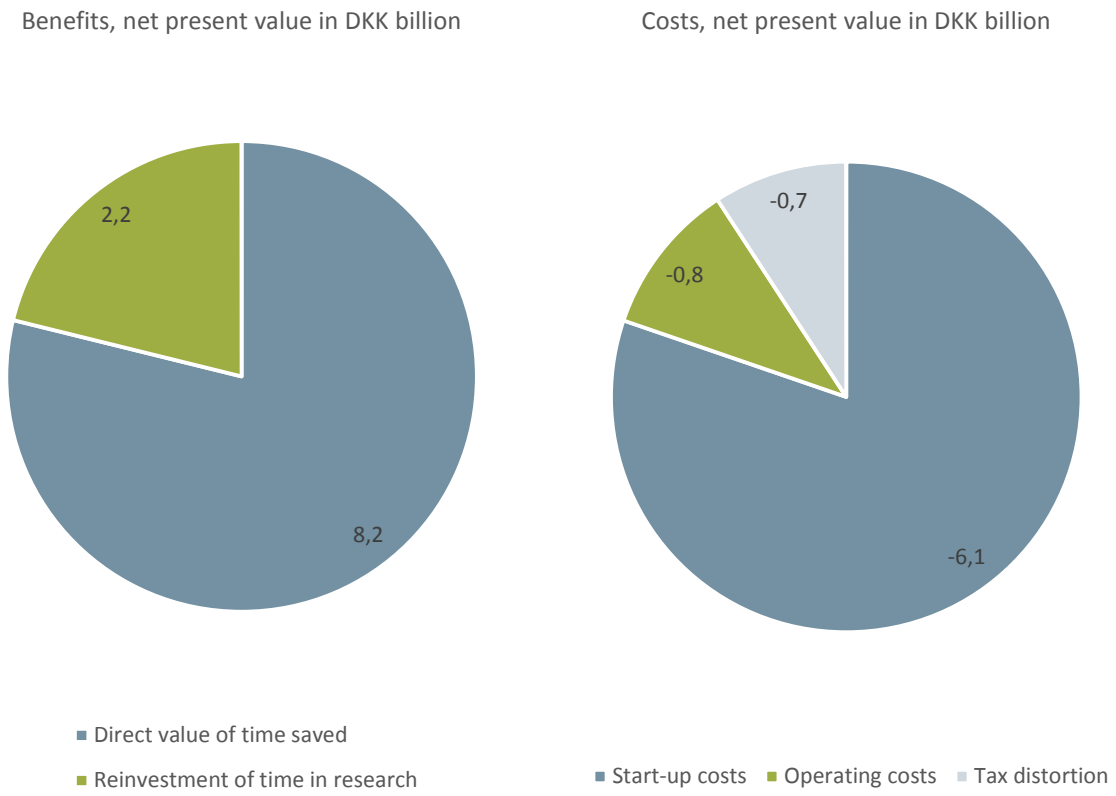
Figure 2.3. Benefits and costs over time



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR. Please note that the values are not discounted.
Source: Own calculations.

The greatest benefit by far of introducing FAIR data in Denmark is the direct value of the time that researchers save, cf. Figure 2.4. In our main scenario we assume that the researchers can save approximately one third of the time they spend on data-related work. The additional return that is achieved by reinvesting the time saved in new research represents a small proportion of the benefits. The greatest costs by far are start-up costs, which make up approx. four fifths of the costs.

Figure 2.4. Breakdown of benefits and costs



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR. The figures are rounded.
Source: Own calculations.

2.3 SENSITIVITY CALCULATIONS

As previously mentioned, our calculations are based on a number of assumptions associated with some uncertainty. In our opinion, however, the result – that the introduction of FAIR data in Denmark will produce a socio-economic benefit – is fairly robust. The reason for this is this result will not change if reasonable changes are made to any of the most uncertain calculation assumptions.

Both the useful life of the FAIR data solution and the time that researchers save on data work are particularly crucial to FAIR data being able to create value in Denmark. If the solution were to have a useful life shorter than 10 years, introducing FAIR in Denmark would be associated with a socio-economic loss. So this means it is important to come up with a sustainable solution. If the time saved is less than 23 per cent of the time that researchers spend on data work, then the introduction of FAIR data will similarly result in a socio-economic loss. This means it is vital to establish a solution that can make the most efficient use of the researchers' working time as possible.

Below we present a number of illustrative calculations of the consequences of changing any of the most critical assumptions in the analysis. Taken as a whole the assumptions cover five areas: 1) time horizon; 2) value of research data; 3) time saved for researchers; 4) reinvestment of time and new research produced; and 5) costs. Sensitivity calculations are presented for each area.

2.3.1 Time horizon

Figure 2.5. Time horizon parameters

- FAIR data implemented over a period of 20 years (source: interviews).
- FAIR data infrastructure has a useful life of 20 years (source: interviews).
- Data have a useful life of 10 years (source: Houghton & Gruen (2014), among others).
- Data depreciate by 10 per cent per annum (source: Houghton & Gruen (2014), among others).
- Benefits and costs are realised after one convex function in the implementation period, i.e. the greatest benefits and costs are realised at the end (source: Houghton & Gruen (2014), among others).

Note: ● High validity ● Medium validity ● Low validity. For a full list of sources, see Chapter 5.

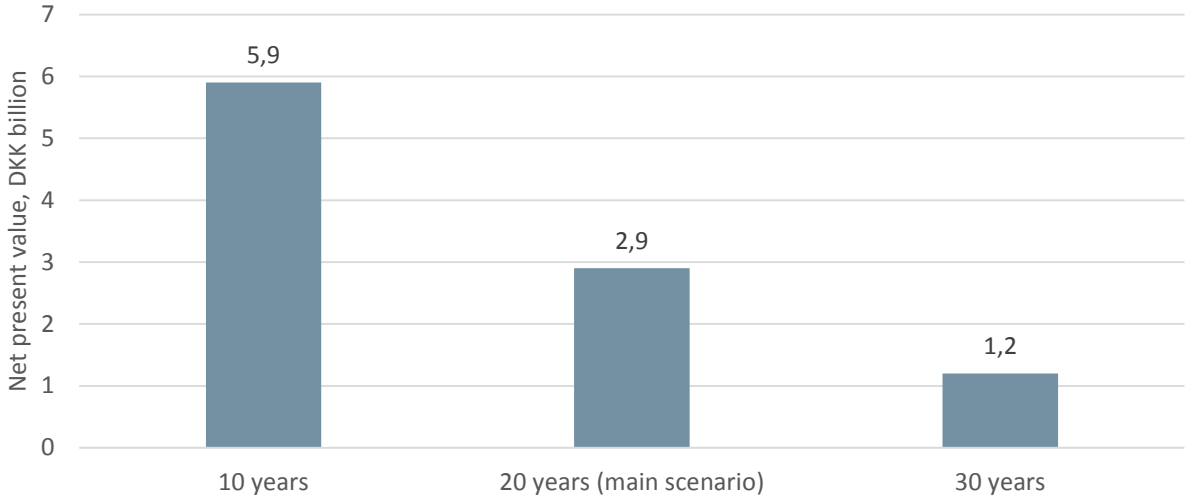
In theoretical literature there are good arguments for the benefits of introducing FAIR data being realised after a convex function. The benefits will continue to accumulate and increase more and more as more data become FAIR. Similarly, it can be argued that the first data to be made FAIR should be those that are easiest to make FAIR. By following this approach, the initial costs will be small but increase as it becomes more difficult to make the last data FAIR.

There is also literature that analyses the useful life and depreciation of public research data. In comparison to the depreciation rates used in the literature, we have applied a relatively high depreciation level. There is somewhat greater uncertainty surrounding the useful life of the data since data today are being generated much faster and in much larger volumes than in the time periods analysed in the literature.

The existing analyses of FAIR research data have only looked at benefits and costs for one year and typically after the FAIR data concept has been fully implemented. In contrast, our analyses of FAIR examine benefits and costs over time and both before and after the FAIR data concept has been fully implemented. This means that there is no basis in the literature on which we can base our assumptions about the implementation period and the useful life of the infrastructure. Our assumption of an implementation period of 20 years is based on the German key stakeholders' early experiences and assessments of the German implementation process, as well as the initial assessments made by the Danish key stakeholders. The conclusion from our interviews in Germany is that full implementation of FAIR will be a long process that takes place at several different speeds given that the federal states' research priorities are not identical. That makes it difficult to estimate the number of years in Germany, but as several people have pointed out this is an implementation process in which the researchers and their communities will slowly become used to following these principles. There are also major generational differences in researchers' attitudes to managing and sharing research data, which is why the belief here is that the implementation only will take off properly once the younger researchers replace the older generation. As a result, an implementation period of 20–30 years is not unrealistic.

Our overall assessment is that FAIR could be implemented more quickly in Denmark than in Germany. That is why we have decided to set a period of 20 years. Factors behind this decision also include the fact that the Danish research community and political system in general are less complex than in Germany, something which suggests that the implementation period could be shorter in Denmark than in Germany. Our sensitivity analyses do, however, show that even with an implementation period of 30 years there would still be a positive socio-economic value from introducing FAIR data in Denmark, cf. figure 2.6.

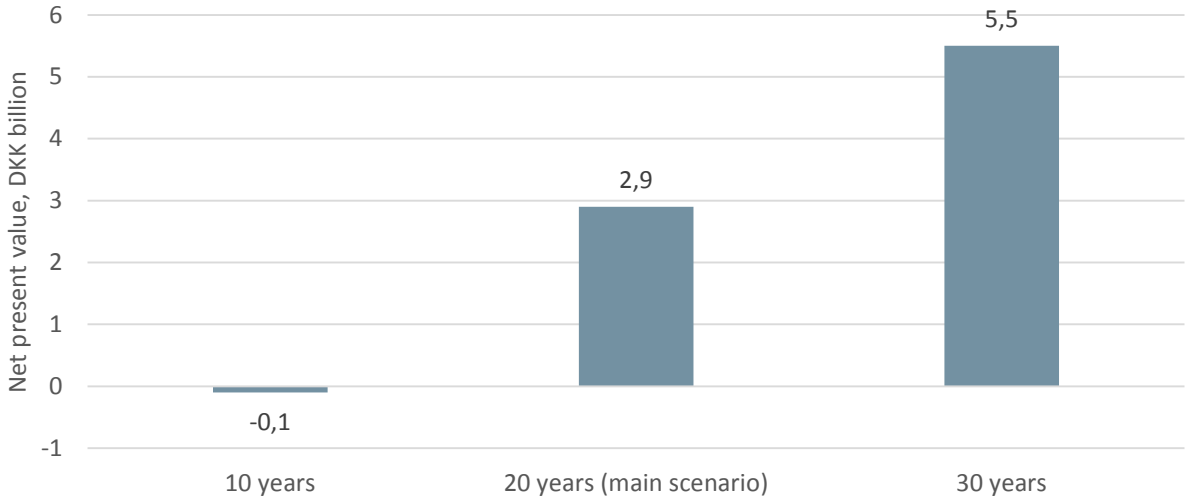
Figure 2.6: Importance of the implementation period



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR.
Source: Own calculations.

In the calculations it is assumed that the infrastructure has a useful life at least as long as the implementation period of 20 years. There is significant uncertainty associated with this assumption. It is likely that parts of the infrastructure, e.g. metadata or user interface, will need to be renewed more quickly while e.g. guidelines, legislation and know-how may have a much longer useful life. Since our analysis includes relatively high operating costs, it is considered realistic to assume that operating costs also include maintenance of infrastructure, including regular updates and any changes to e.g. metadata and the user interface. A useful life of 20 years is therefore considered realistic. Our sensitivity calculations show that the socio-economic value of FAIR data is positive, even with the infrastructure having a useful life of 11 years, cf. figure **2.7**.

Figure 2.7. Importance of the useful life of the infrastructure



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR.
Source: Own calculations.

2.3.2 The value of research data

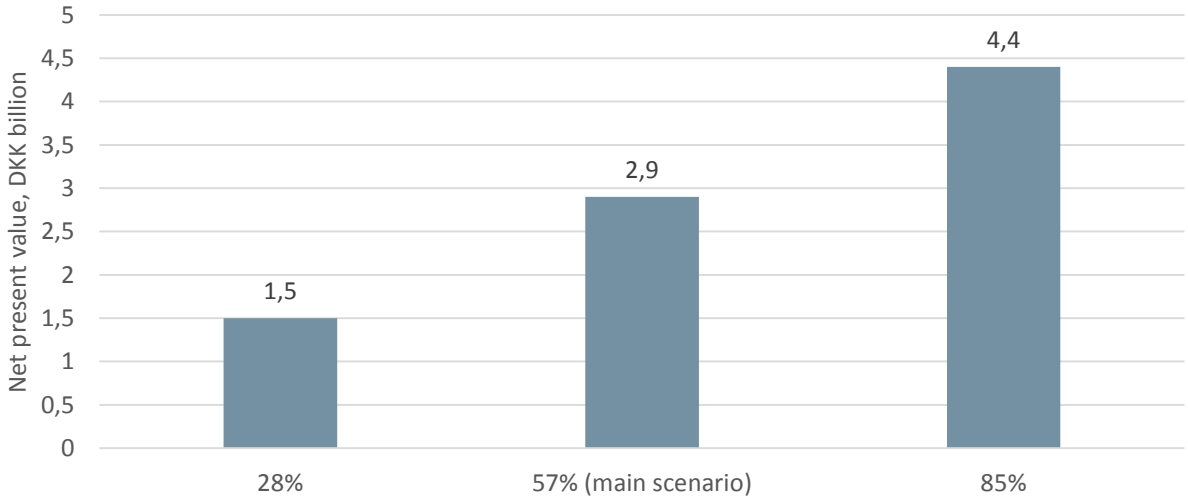
Figure 2.8. Parameters for estimating the value of research data in Denmark

- Proportion of researchers' time spent on research and knowledge (source: DEA (2015)).
- Proportion of research time spent on data (source: Beagrie & Houghton (2016), among others).

Note: ● High validity ● Medium validity ● Low validity. For a full list of sources, see Chapter 5.

Researchers devote their time to lots of activities other than research and knowledge creation. In 2015 the think tank DEA conducted a study of Danish researchers which found that they spent approx. 48 per cent of their time on research. Since this study is relatively recent and deals with Danish researchers, we deem it to have high validity. There is no empirical evidence of the time that Danish researchers spend on data work. We have therefore based our assumptions about the time spent by researchers on data work on the international literature. The information in the literature was typically obtained through questionnaires given to researchers who use data centres. The estimates from the existing literature are therefore considered to be upper bound estimates in relation to the time that the average researcher spends on data work. In our calculations both the benefits and costs vary according to the time that researchers devote to data work. When the researchers spend more time on data work they generate more data. This will lead not only to greater benefits from introducing FAIR data but also to higher costs. In our calculations it is assumed that both benefits and costs constitute a fixed proportion of the value of research data in a given year. This means that, relatively speaking, there is an equal rise in benefits and costs if researchers spend more or less time on data work. In other words, the costs will never exceed the benefits regardless of how much time the researchers devote to data work. The assumption does, however, affect the size of the socio-economic value of making data FAIR, cf. figure 2.9.

Figure 2.9. Importance of the proportion of research time spent on data



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR. The figure shows the net present value for a proportion of time spent by researchers on data work 50 per cent lower or higher than what is used in the main scenario.
Source: Own calculations.

2.3.3 Time saved by researchers

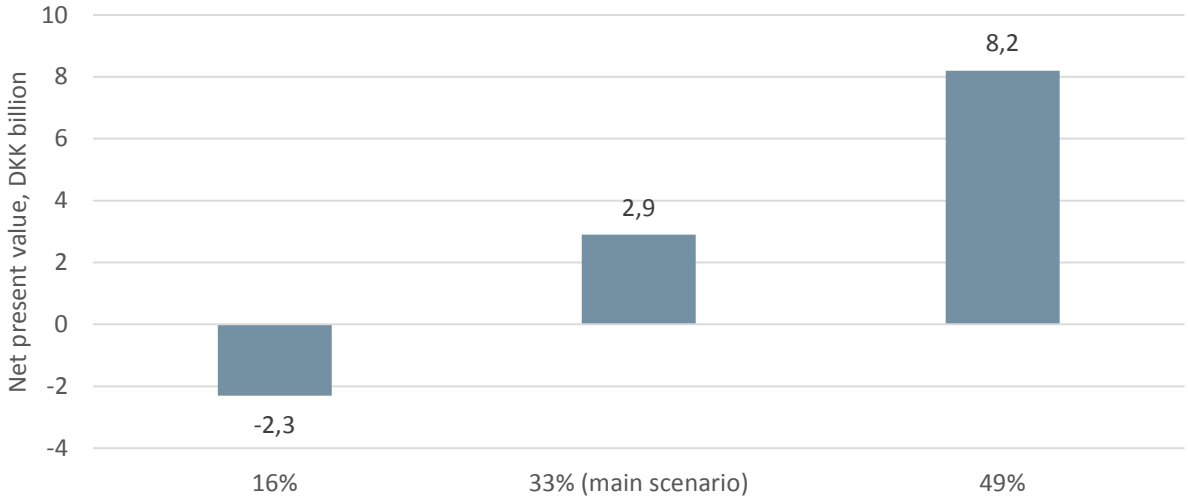
Figure 2.10. Parameters for estimating the value of time saved for researchers

● Efficiency gain, proportion of time that can be saved on data work (source: Sullivan et al., among others).

Note: ● High validity ● Medium validity ● Low validity. For a full list of sources, see Chapter 5.

The proportion of time spent on data work that can be saved when data are made FAIR has been calculated in several international studies. These studies are based on questionnaires given to users of FAIR data centres. The estimates are therefore considered to be upper bound estimates in relation to the time that the average researcher can save when data are made FAIR. We have consequently decided to use the lowest estimates for the time saving found in the literature, and in our main scenario we thus apply an assumption that researchers can save approximately 1/3 of their time when data are made FAIR. The literature contains estimates of up to 46 per cent. Our sensitivity calculations show that there is a positive socio-economic value from introducing FAIR in Denmark provided that the time saved/efficiency gain is 23 per cent or higher. In other words, the result is relatively sensitive to the uncertainty associated with the efficiency gain.

Figure 2.11. Importance of proportion of time that can be saved on data work



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR. The figure shows the net present value for a 50 per cent lower or higher time saving than what is used in the main scenario.
Source: Own calculations.

2.3.4 Reinvestment of time and new research produced

Figure 2.12. Parameters for estimating the value of new research

- Return on new public research (source: DEA (2017), Universities Denmark (2016)).
- Only some of the return realised in Denmark (source: Houghton & Gruen (2014), among others).
- No new research produced (conservative assumption).

Note: ● High validity ● Medium validity ● Low validity. For a full list of sources, see Chapter 5.

The literature on the return on research and development is extensive. In recent years several literature studies have been carried out in Denmark; their assessment was that the return is in the order of 20–40 per cent.[15] It should be noted though that a great deal of the literature does focus on private investments in research and development. We have opted to go with 20 per cent, which is at the lower end of the range. Given there is some uncertainty about the return on public research and development, the estimate does not, however, have high validity. There are also several studies that point to the return on research being achieved both locally and internationally. There are no studies of this for Danish data, and we have therefore chosen to base our calculations on an estimate at the lower end of the range identified by the international studies.

[15] Sources: DEA (2017), Universities Denmark (2016)

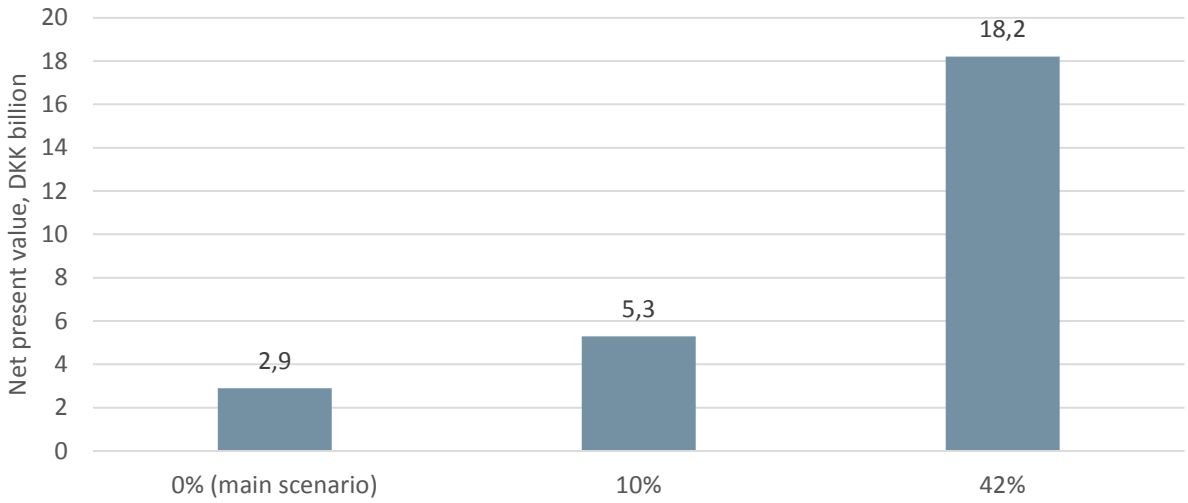
The value of reinvestment of time accounts for a relatively small proportion of the benefits in the main scenario. Even without any extra return on the time saved by researchers due to data being FAIR, there would still be a positive socio-economic value from introducing FAIR data in Denmark.

In our main scenario we have, as previously mentioned, disregarded the value of new research produced when data are made FAIR. This is mainly due to the uncertainty over the extent to which new research will simply displace existing research. Yet there may be reason to believe that the new research will have a positive socio-economic value. One such reason may be that researchers will continue to do much of the research they used to do before data became FAIR (corresponding to a rise in their total research activity), but it may also be because the return on the new research produced by FAIR is higher than the return on the research previously carried out by the researchers. If we assume that researchers are rational, then they will pick the research that generates the greatest return.

In Houghton & Gruen (2014) it is assumed that new research does not displace existing research. Consequently, they use the proportion of research that would or could not be produced without FAIR research data to estimate the value of the new research. This proportion is based on international literature and questionnaires given to users of FAIR data centres. The outcome of this approach is that the introduction of FAIR research data will lead to a marked increase in total research activities. If we take the same approach as in Houghton & Gruen (2014) and base calculations on the lowest estimates in the literature of how great a proportion of the research would or could not be produced without FAIR data (42 per cent), then the socio-economic benefit of introducing FAIR data in Denmark will be approx. DKK 18 billion. This example is a clear upper bound estimate and illustrates first and foremost the possible potential of introducing FAIR data in Denmark.

If a more realistic assumption is applied, i.e. that 10 per cent new research is produced or that the new research will generate an approx. 10 per cent greater return, the socio-economic value will be DKK 5.3 billion, cf. figure 2.13.

Figure 2.13. Importance of proportion of new research produced as a result of FAIR data



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR.
Source: Own calculations.

2.3.5 Costs

Figure 2.14: Cost parameters

● Start-up costs on average make up of 24 per cent of the annual data value in the implementation period (source: Beagrie et al. (2010)).

● Operating costs on average make up of 3 per cent of the annual data value (source: Beagrie et al. (2010), among others).

Note: ● High validity ● Medium validity ● Low validity. For a full list of sources, see Chapter 5.

There are few studies in the literature that deal with the costs associated with introducing FAIR data or other open data solutions. They largely focus on standalone initiatives and typically assume that the decision-makers are able to a large extent to determine the scope of their investment for themselves. There is thus a great deal of uncertainty involved in working out the costs associated with the introduction of FAIR data in Denmark. The operating costs are less uncertain because a number of initiatives already exist that ascertain these. In our calculations, we have chosen to set operating costs at 3 per cent of the total wage costs for data in the individual years,[16] a figure which is significantly higher than in Houghton & Gruen (2014), which only uses 1.5 per cent.

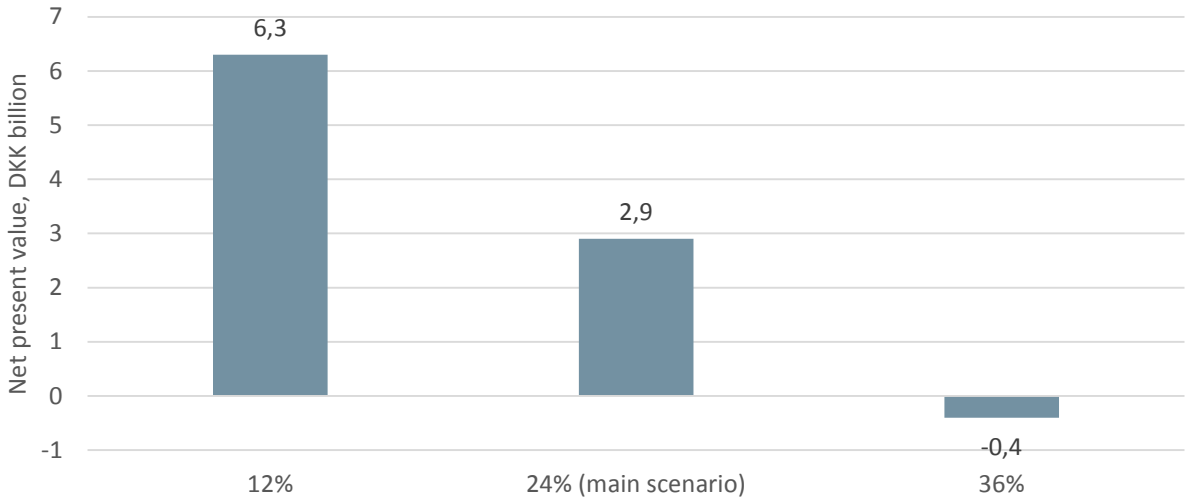
The start-up costs are modelled on the basis of a single case from Oxford University where they amounted to 24 per cent of the total data value. There is great uncertainty surrounding the validity of this estimate because the period and the project in question differ from an introduction of FAIR data in Denmark. We believe, however, that the estimate is an upper bound estimate, cf. Chapter 5, on the basis of the interviews conducted and owing to the potential savings that new technology, e.g. in the form of Artificial Intelligence, may generate compared to the project in Oxford. Our results are very sensitive to start-up costs, cf.

figure 2.15.

[16] These costs are estimated using the same method as for the value of data, i.e. the time that researchers spend on generating data and the wage costs for this.

Despite this, there is still a positive socio-economic value for start-up costs – up to 35 per cent on average over the period.

Figure 2.15. Importance of start-up costs



Note: It is assumed that 75 per cent of all research data in Denmark will be made FAIR. The figure shows the net present value for a 50 per cent lower or higher proportion than what is used in the main scenario.
Source: Own calculations.

3. Opportunities for and barriers to introducing FAIR data in Denmark

The purpose of this chapter is to analyse the barriers and opportunities in implementing FAIR data in a Danish context, as well as investigating which policy, organisational, technical and cultural measures can be initiated to roll out FAIR data in Denmark.

3.1 EVALUATION OF THE FAIR PRINCIPLES

As previously described, FAIR is based on the following four principles – data must be Findable, Accessible, Interoperable and Reusable. These four principles are closely related, yet in technical terms are independent of each other. They can be implemented in various combinations and in different degrees or stages. A full implementation of the four principles of FAIR is more of an ideal in a continuum where data can be made more or less FAIR[17]. The fact that this is a question of principles – not set requirements – which can be followed to different degrees is also underlined by the Chairman of DM Forum.

> *"There are thus several degrees of FAIRness, and DANS [Data Archiving and Networked Services] in the Netherlands are currently working on developing a ranking for the individual principles."* (Anders Sparre Conrad, Chairman of DM Forum)

Our interviews with the Danish key stakeholders, including universities, research libraries and the policy authorities, show that there is general backing for the FAIR principles in Denmark. Several emphasise that it is a matter of sensible principles, something which it is difficult to disagree with. Having said that, it is difficult for most representatives from research institutions to find out exactly what is underneath the individual principles. To commit to individual principles properly it is necessary to make them more practice-based and articulate some practical implications; this is also the underlying idea behind FAIR.

Even if it is difficult to commit tangibly to the individual principles, the following sections still summarise some of the key assessment points – partly on the basis of the interviews conducted in Denmark and partly based on the experiences in Germany.

3.1.1 Findability

To comply with the first FAIR principle, data must be findable. This requires good metadata that describe data. At the same time, it is also important for machines to be able to find data, as data will have typically spread across boundaries – both geographical and technical – and by virtue of the fact that in some cases the volumes of data are too large for humans to handle them. To meet this principle, the data should be assigned a unique and persistent identifier, also called a PID, and be stored in a repository with an extra layer of information about organisation, structure, metadating, etc.

[17] Barend Mons, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos & Mark D. Wilkinson (2017). Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud.

All the Danish stakeholders in this area share the assessment that data to a certain extent are already findable and that from a technical perspective the principle of findability is relatively easy to meet. However, our analysis also shows that 14 of the 20[18] international and national data repositories with Danish participation do not use PIDs. This demonstrates that most of the data storage used by Danish researchers is nowhere near meeting the criterion to be Findable and is therefore far from qualifying as FAIR data.

There are already examples of local Danish repositories, including ERDA[19] and Dataverse[20], as well as national repositories, such as LOAR[21] at the Royal Danish Library and the Danish nodes in international repositories such as CLARIN-DK[22]. Globally speaking there are also several generic and discipline-specific repositories, e.g. ELIXIR[23] and CESSDA[24]. Yet the problem here is actually being connected to an efficient search engine that requires constant maintenance and ensuring that the researchers have the right skills and incentives for data registration, particularly metadating and sharing of large volumes of digital research datasets. These are absolutely key challenges to the introduction of FAIR data and which are covered in more detail later on in this chapter.

3.1.2 Accessibility

To comply with the second FAIR principle, data must be accessible given that they are stored for the purpose of long-term preservation. There must also be well-defined licensing rules and it should be easy to retrieve data. It is moreover important here to point out that the *accessible* criterion of FAIR is not synonymous with all forms of data having to be open and accessible to the public. Data which, for a variety of reasons, are deemed to compromise sensitive personal data, national security, competitive conditions or other similar important matters are covered by special legislation that lays down a dedicated framework for data accessibility; these data may only be accessed as agreed in a separate contract/licence. This is set out explicitly in Horizon 2020's underlying guideline on FAIR data management, i.e. *"as open as possible, as closed as necessary."*[25] But it is important to emphasise that even if data are not accessible, they must be findable, understood in the sense that metadata on the material can be found but for various reasons not always the actual raw data.

The universities and other research institutions are generally very aware of the principle of *accessibility*. It is expected that requirements in instruments such as the EU's new General Data Protection Regulation will mean that in future more datasets must be security-certified to merit inclusion in any research collaboration and thus will also require more resources to meet the principle of data being accessible. Similarly, the

[18] Re3data.org
[19] ERDA is a storage, sharing and archiving system provided by UCPH to its employees and students. In the first instance, access is purely for the Faculty of Science, but in the longer term may be extended to the whole university.
[20] The Royal Danish Library and Copenhagen University Library and Information Service have created the KUBIS Dataverse Network, a digital archive for primary research data.
[21] LOAR (Library Open Access Repository) is an open data repository, established in 2016 as a service for storing and providing access to Danish research data.
[22] CLARIN-DK is the Danish contribution to a European research infrastructure for the field of humanities. CLARIN-DK controls a platform on which researchers can access the language-based materials and tools gradually being deposited by researchers from different universities and institutions.
[23] ELIXIR is the European research infrastructure for bioinformatics.
[24] CESSDA stands for Consortium for Social Science Data Archives and provides cross-border access to social science data.
[25] European Commission (2016), H2020 Programme – Guidelines on FAIR Data Management in Horizon 2020.

opportunities for and barriers to *access* to relevant research data as an element of a Danish ecosystem for digital research data are expanded on later in this chapter.

3.1.3 Interoperability

To comply with the third FAIR principle of *interoperability*, it is crucial for there to be common guidelines on describing metadata and for an open and well-defined vocabulary to be used. This facilitates both the use of data across research fields and cross-system data exchange. Full interoperability means that data can be found, accessed and used across separate physical data repositories and are not dependent on a particular physical and technical platform.

Denmark does not currently have any initiatives in place that meet the third FAIR principle of interoperability, but in Germany thanks to e.g. the GO FAIR initiative it has been possible to break with the previous practice of having multiple independent repositories and instead try to achieve more interdisciplinary coordination to avoid solutions that cannot be combined. The GO BUILD element within GO FAIR addresses e.g. the need for datasets that use harmonised standards and which make it possible for all researchers to access data across subject-specific repositories.

Many interviewees considered interoperability to be one of the greatest challenges associated with making data FAIR. In Germany it was decided not to focus on a consolidated data solution but rather to promote coordination of the already established German data repositories as well as greater international cooperation such that the technical standards used for preservation of research data apply not only to Germany but internationally, including through RDA (Research Data Alliance) and EUDAT (Research Data Services, Expertise & Technology Solutions).

Just like Germany, Denmark is a participant in numerous initiatives including RDA and EUDAT. The Danish representative in RDA explains that the principle of interoperable research data is one of the most difficult principles to meet because it requires data to have the capacity to communicate across research fields and identify a common language. Having said that, there are some international projects in progress under the auspices of EUDAT that are testing the interoperability of the metadata level.

> *"There are many format challenges inherent in making data interoperable. Right now we are testing the metadata level under the auspices of EUDAT. Far ahead in the future we may perhaps be able to make them interoperable at raw data level, but this is very difficult because it is layer-to-layer information."* (Lene Krøl Andersen, Head of eScience Competence Center, DeIC)

3.1.4 Reusability

It must be possible to reuse data in future research projects and then process these data further to comply with the fourth and last FAIR principle. One of the key prerequisites for data being reused in other research projects is that data are released using a clear and accessible data licence.

There are a number of legal challenges, including defining rights of use for research data, that must be addressed for it to be possible to reuse data. These rights must be set out clearly because ambiguities in copyright can limit data reuse. The more automatic the process to find stored data becomes, the more important it is to also have a clear and accessible data licence.

Unsurprisingly, a common thread in the interviews we conducted was that the reusability principle is one of the most difficult principles to meet. As, for example, the Chairman of DEFF explains, this is due to the fact that at present no credit is given for data sharing by researchers.

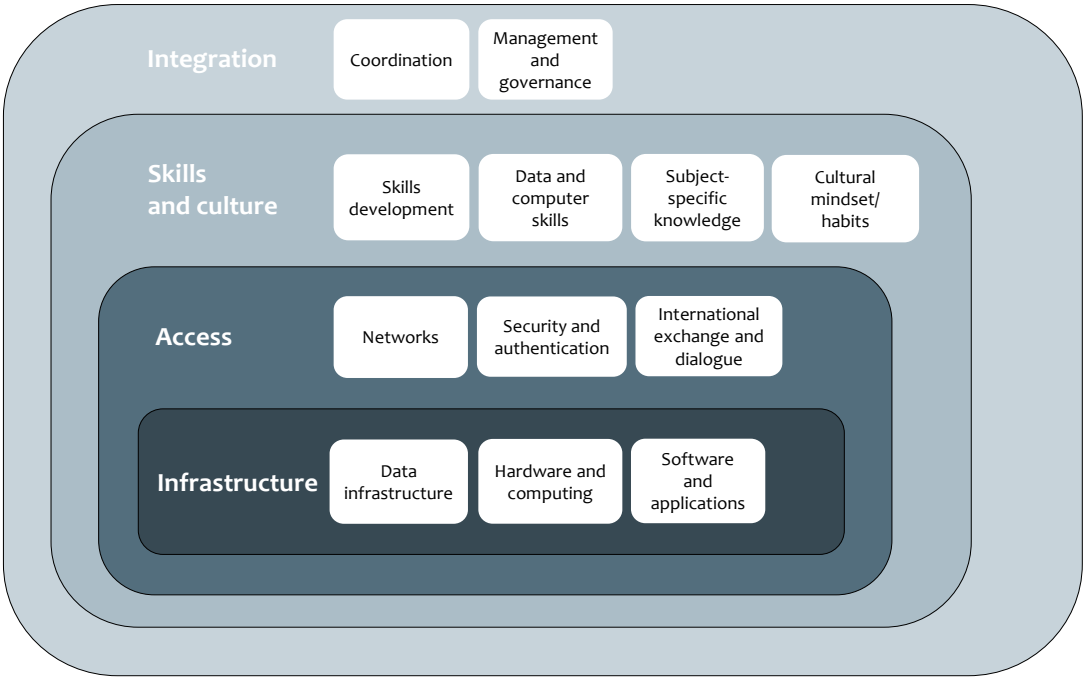
> "*We are a little doubtful about the final principle of reusability. This may be quite difficult to meet as the researchers do not get credit for sharing data. It is necessary to take a look at incentive structures.*" (Professor Børge Obel, Centre Director at Aarhus University and Chairman of DEFF)

3.2 ELEMENTS OF THE ECOSYSTEM

The Danish ecosystem for digital research data is complex and consists of several distinct elements that need to interact with each other, cf. Figure 3.1 below in which the main elements of the ecosystem are *integration*, *skills and culture*, *access* and *infrastructure*.

For each element there are stakeholders and actors with different requirements, interests and needs. A successful implementation of FAIR data in Denmark depends on many factors. One absolutely key point is that the international connection and integration is essential, since there is no sense in having an isolated, national FAIR data solution. It is also key for many of the stakeholders to see the advantages of the FAIR principles. For example, many of the researchers can see that in the longer term they will be able to improve the quality of their research and save resources on data collection by sharing and reusing each other's data and interpretations. They also realise that the research librarians can play a role in the research processes and help to catalogue the digital data systematically. And that the professional data providers can offer services and supply data analysis and processing services designed to support good digital research practice, not to mention that public research funding, which is increasingly being influenced by an international Open Science agenda, can help to support better research and new discoveries through requirements for long-term data management and data sharing in line with the FAIR principles.

Figure 3.1 Elements of the Danish ecosystem for digital research data



Source: Oxford Research inspired by RCUK e Infrastructure (2014): E-infrastructure Roadmap

The basic task of the ecosystem is to balance the elements and create coherence in such a way that the efforts are coordinated and not standalone. Specifically this means e.g. that focus on further training of researchers to manage and share digital research data should be seen in the context of establishing sufficient numbers of data supporters at the universities and whether there is sufficient data capacity and storage space to store data.

The *integration* of the elements of the ecosystem should establish exactly this context in the system and one of the things it requires is centralised, national coordination that can produce a consistent and integrated system across environments and actors with an eye towards the activities taking place locally, nationally and internationally.

The second main element of the ecosystem is *skills and culture*. The point is that even if the technical solutions and infrastructure are in place, it will not be possible to harvest the potential of FAIR data without possessing the right skills and without a cultural mindset that supports the use of FAIR data. Firstly, the correct management and sharing of digital research data requires the researchers to possess the necessary eScience skills. These must be supplemented and supported by data stewards, eScience experts and research librarians who must to a greater extent take on the role of 'co-creators' in the research process. Secondly, the cultural dimension requires the researchers and institutions around them to see the value of research data and not just of published research articles.

Access is the third element. It requires a research network that functions properly and the researchers to have access to relevant research data via a well-known and simple process. First and foremost it is a question of gaining control of the purely technical network solutions and of linking databases to research networks. In addition to this, there is a great deal of coordination work involved within the individual fields of research relating to cross-border knowledge exchange on the development of metastandards, documentation requirements and data management.

The *infrastructure* is the final element and can be seen as the backbone of the ecosystem since it enables the storage, management, processing and dissemination of digital research data. Infrastructure can exist locally at the individual institutions, nationally in e.g. national data archives and globally in international data centres. Again the point here is that infrastructure cannot stand alone because investments in and the establishment of the requisite specialised infrastructures merely enable the storage, management, processing and dissemination of research data; behavioural changes among the researchers are required in order for research data to actually be shared.

3.2.1 Integration

Integrating the elements of the digital ecosystem requires clear policy prioritisation and a centralised effort focusing on the FAIR principles. That is why it is important to have a unifying actor in place which can set the overall objectives and frameworks, e.g. by developing a strategy for the area, and also be responsible for interdisciplinary coordination and helping to create stakeholder consensus in the system. It is absolutely crucial for researchers, universities and other research institutions, research councils, foundations and libraries to work towards the same goals in relation to e.g. data management policies, courses in data management, data steward training and also incentive and merit structures. Integration must therefore ensure that FAIR data are implemented in the most appropriate and resource-saving manner.

3.2.1.1 A common, coordinated and national effort

Denmark does not currently have a national strategy, policy or common guidelines in place to support the FAIR principles and put them into practice. Coordination in Denmark is based on the activities initiated by DeIC, DEFF and DM Forum. DeIC and DEFF worked together to come up with the National Strategy for Research Data Management 2015–2018. The strategy is widely embedded and has been sent out for consultation with the main stakeholders in Denmark. Its purpose is to support a common Danish trend for collection, security, dissemination and reuse of research data. One recommendation in the strategy is to establish a national forum for cooperation on research data management which, with financial support from DEFF, will promote the key strategic recommendations with a focus on policies, incentivisation, infrastructure, skills development, research grants and management structure. On the basis of the recommendations set out in the strategy, DM Forum has technically speaking been set up as a three-year pilot project and will serve as the Danish focal point for the practical exchange of experiences and collaboration on the development of expertise and infrastructures; it will also coordinate the Danish effort in connection with international activities. DM Forum currently supports research data management in general and not FAIR specifically. However, DM Forum does emphasise that the organisation is open to the FAIR principles and is keen to help coordinate and support their implementation.

The work with open data is, however, still relatively uncoordinated because individual institutions, including universities and research libraries, each draw up their own policies and appoint groups that work with research data management specifically. To take just one example, DTU, which has achieved the most progress with FAIR data in Denmark, set up a data management group (DM Group) in 2015 that the researchers can consult on data management issues. At the same time, DM Group is in the process of de-

veloping policies and practices that can then be incorporated into the data management work. The impression we have gained from the interviews is that the universities do not have a complete overview of what is going on in Denmark in this area. Despite this, there are examples of collaboration between the universities and of them exchanging knowledge and inspiration. One aspect of SDU's Open Science Policy, which is under development, is that it is inspired by DTU's policy in this area; SDU is currently building upon an infrastructure that AAU has moved away from. The agreement is that whatever is being built at SDU will also be built at AAU.

Germany, which has officially opted to embrace the FAIR principles, has launched the GO FAIR initiative as the focal point for implementing FAIR in a German context. Thus far the initial German experiences show that the ministerial backing and earmarked funding for the GO FAIR initiative have been a significant driver for putting FAIR data on the agenda in Germany. The purpose of the GO FAIR initiative is to create widespread support among the key stakeholders as well as focusing on the potential of FAIR by means of specific best practice examples that illustrate how the FAIR principles can be operationalised; one way of doing so is through projects such as GeRDI[26] and RADAR[27].

3.2.1.2 The connection between local, national and international

One key point is that an isolated, national solution for FAIR data does not make sense. The international connection and integration is critical to the researchers and institutions seeing an advantage in applying the FAIR principles and changing their practices. There are several reasons for this. First and foremost, it can be said that the nature of research is inherently international. This means that the researchers should be able to share and reuse data with their peers at an international level and that it is not good enough if researchers are only able to share and reuse data nationally in Denmark.

> *"The starting point for a solution is internationality. The implementation basis for FAIR must be the researchers who work internationally. Within their own research disciplines researchers must exchange knowledge, practices and experiences, which can then be supported nationally."* (Henrik Pedersen, Dean of the Technical Faculty of IT and Design, AAU)

Secondly, it is important to keep up with international developments and coordinate the Danish effort in connection with international activities to serve as a guarantee that no local and national solutions will be

[26] The long-term aim of GeRDI (Generic Research Data Infrastructure) is to make it possible for all German researchers to store, share and reuse research data across fields of research. The first phase of the project will see the merger of three data centres. This should create the pilot case for roll-out in Germany being possible and serve as a model for the future German research data infrastructure.

[27] RADAR (Repository Research Data) is about collecting data in a data centre to which the various different members have access. The project is interdisciplinary and its participants come from universities in different German states. The project is organised so that there is a local coordinator at each institution who is responsible for supporting the local researchers in data sharing and data management at their own institution. The project supports 'the long tail' within research data, which is those data that fall outside of the big data domain, i.e. first and foremost small datasets. In terms of FAIRness, the assessment is that the project has achieved the criteria of Findability and Accessibility but is far from meeting Interoperability and Reusability.

developed if e.g. international solutions already exist that can be brought into use. Work is currently underway to make sure that the Danish effort is in keeping with the work in key international cooperation forums such as the Research Data Alliance (RDA), Knowledge Exchange, e-Infrastructure Reflection Group (e-IRG), European Data Infrastructure (EUDAT) and Nordic e-Infrastructure Collaboration (NeIC). As a supplement to ensure a connection between the local/national and international initiatives, the GO FAIR initiative could be viewed as a possible platform for international cooperation. According to Barend Mons, one of the GO FAIR organisers, they are in the process of setting up working groups focusing on the implementation of the FAIR principles.

> *"I would recommend that Denmark join the GO FAIR implementation networks which are multi-topical and user case driven – we will for example jointly develop requirements for good data management. Both institutions, funding agencies and the Ministry can join the implementation networks."* (Barend Mons, Professor at Leiden University)

3.2.1.3 Sustainable funding

Implementation of the FAIR principles necessitates investments in the establishment of new structures, processes and systems. There will also be a long transition and operating phase to get the principles of research practice operationalised and integrated. Investment is needed in many different areas, including to create the institutional frameworks for FAIR, infrastructure investments, communication and coordination processes in research environments, the reorganisation of existing services, training and further training of researchers.

Generally speaking, the Danish stakeholders have a positive view of the potential that can be harvested by implementing FAIR data in Denmark, but several of them are uncertain of the funding model and whether the costs must be allocated to the researcher or the research institution.

> *"The funding is unclear. The costs associated with maintaining data are huge. My experience from the foundation world is that we receive a large number of applications for funds for data maintenance. But when is it actually worth maintaining data? And does this cause barriers to generating new data?"* (Professor Jens Oddershede, Chairman of the Danish Council for Research and Innovation Policy (DFiR))

In Germany, several of the stakeholders, including Dr Stefan Winkler-Nees from Deutsche Forschungsgemeinschaft (DFG), points out that a major challenge in Germany is sustainable funding. Many of the Open Science projects are funded with project funds, meaning there is a risk of them foundering when this money runs out. Consequently there is a need for more long-term institutional funding, although new business models also need to be devised and followed. Among other outcomes, the German RADAR project was responsible for developing a business model in which researchers and research institutions must pay to participate.

3.2.1.4 Requirements for policies and practices surrounding research data

Data Management Plans (DMPs) are a vital tool for promoting and strengthening the one-way direction of data management and data sharing. A DMP is a document that describes in a structured manner how

the researcher intends to manage their research data with regard to retrieval, sharing, ethical considerations and not least data storage etc. It is a simple text document, typically in PDF format, and it is the researcher themselves who determines what it should say.

The German experiences and the requirements of the EU's Horizon 2020 research programme that researchers who receive grants from the programme must submit a DMP together with their application demonstrate there is now greater focus among the researchers and in the institutions' policies and practices. DFG in Germany stipulates that a DMP must be prepared as part of the application process, and management at several of the Danish universities believe that the EU's Horizon 2020 research programme's requirement for researchers who receive grants from the programme to submit a DMP together with their application has led to greater focus among the researchers and in the institutions' policies and practices. As noted in several of the interviews we conducted, the requirements for DMPs are not, however, synonymous with data from funded research projects actually being stored in repositories that make data accessible to other research projects.

A number of the informants emphasised that the lack of a requirement to prepare DMPs at the institutions is a real challenge in Denmark. DM Forum's overall vision is that by June 2018 Denmark will have a national, coherent set of data management (DM) policies at its universities which 1) are backed by research councils, ministries and repositories and are open to other research institutions; 2) ensure that the researcher gets help planning and executing data management throughout the data life cycle; 3) are underpinned both locally and internationally; and 4) establish models for funding and distribution of the costs of DM between the stakeholders. In 2014 a Danish Code of Conduct for Research Integrity was also developed. Feedback from the universities indicates that the Code of Conduct has helped to stimulate considerations and thoughts about research data management and that their policies have taken inspiration from this Code of Conduct. At the same time, the interviews with several different universities and researchers demonstrate that it has not yet led to definite changes in their research practice, since this will require a broader effort in several areas, including a change in the researchers' incentive structure, their skills development and more investments in infrastructure, etc.

3.2.2 Skills and culture

Skills and culture are essential prerequisites for achieving FAIR data in Denmark. Since it is vital for the researchers to apply the principles in their work and act as allies in the operationalisation process, the starting point for introducing FAIR must be the researchers themselves.

Firstly, skills play an absolutely essential role because the correct management and sharing of digital research data requires the researchers to possess the necessary eScience skills. These must be supplemented and supported by additional specialist data stewards, eScience experts and research librarians who must to a greater extent take on the role of 'co-creators' in the research process and possess knowledge of data management with insight into the research domain. Additional supplementary data expertise is also required to support the development and maintenance of data infrastructure.

Secondly, there is the cultural dimension, since both the researchers and the institutions around them need to change their mindset and see the value of research data and not just of publications from now on. Barend Mons predicts that although researchers currently publish articles with an appendix outlining the data,

in future data will be the primary publication to which authors can then publish a supplementary article[28]. Many measures are needed to support the culture change; the most important of these is probably an adjustment to the incentive and merit structure, which at present is centred on the publication of scientific research articles. A change to the existing merit practice should in that case take place in an international context by virtue of the international nature of research.

3.2.2.1 Developing researchers' data skills and data support

The current situation in Denmark shows that there is not a systematic approach to skills development and training of researchers with a view to providing them with the requisite tools for managing and working with FAIR data. The data skills of researchers have not been on the agenda at Danish universities, and a common theme in the interviews was that researchers do not have adequate data management skills. Since several of the universities have not coordinated their efforts in this area, responsibility has been at faculty and departmental level and often rested with the individual researcher. This has been problematic because the researchers rarely have the requisite knowledge or insight. As a consequence, data are often stored locally and not shared or managed consistently.

> *"Virtually all researchers have no training at all in data management. But there are some universities, like UCPH and AU, which have started to offer courses in this subject."* (Deputy Director General Bjarne Andersen, Head of IT Development and Infrastructure at the Royal Danish Library)

Another focus area that it is important to support is the establishment and expansion of data supporter roles at the universities. The purpose of creating such roles is that it means the researchers can receive advice on the use of infrastructures and IT tools, metadating and issues relating to the organisation of information, as well as on legislation and legal issues.

> *"I fear that when a researcher sees how much time and how many resources it takes to clean and document data it will all come to a halt. They will ask themselves: 'What's in it for me?' That's why they need to have data stewards who can help them."* (Nikolaj Helm-Petersen, Deputy Head of Universities Denmark)

At present Danish universities are watching how support roles are evolving but are not making any major, specific investments. According to Barend Mons, Chairman of the EOSC High Level Expert Group, the Expert Group has estimated there is a need for something like one data steward per 20 researchers if the potential of the EOSC is to be exploited fully.

In addition to this, there is a need for a closer link between research libraries, data archives and universities, and in this context a new library science degree course that combines traditional librarian training with computer science may be a solution. John Renner, Dean of KU Science, believes that a new library

[28] DeIC (20/12/2016): Data will replace the scientific article, link to news article in Danish: https://www.deic.dk/da/news-2016-12-20_Barend_Mons

science degree or a data steward training course is needed to cover the overlap between a traditional librarian and a computer scientist. An analysis of research data and Open Access commissioned by DEFF concludes that even a highly experienced librarian who knows how to work with researchers, systems, research registration, metadata, IT, etc., finds it very difficult to understand a research data area because it is outside their comfort zone and each case has its very own content and purpose and its very own background and special management requirements[29].

3.2.2.2 Time and resources

The fundamental issue is that it takes time and resources for the researchers to migrate from their previous research practice to a greater focus on data sharing and data management. This is due to the fact that data sharing involves strict formatting, documentation and maintenance requirements, a factor which then takes time and resources from something else.

> *"I think that the cost is greatest in the human resources required to support the entire life cycle of the data and the documentation it entails. The operating costs are the ones that are relatively invisible, because they can become part of your normal work."* (Professor Børge Obel, Centre Director, DEFF).

A fundamental challenge here is that the current merit system does not consider researchers who use resources to collect and maintain data, and their time and costs are not acknowledged. Placing more emphasis on the sharing and use of research data in the credit is naturally the most effective way to change researchers' behaviour. In all the interviews with the Danish universities a theme that kept appearing was that this system requires an adjustment if it is to change the researchers' approach to research data. The same thing is apparent from the German experiences, where the interviewees also point out the restrictive effect of this system. It is an international challenge that demands internationally coordinated solutions and interdisciplinary collaboration and will only be resolved in the long term.

3.2.2.3 Researcher mindset

The experiences of researchers with data sharing, their data sharing practices and their motivation to share data vary depending on the research discipline. That is the conclusion reached by Knowledge Exchange in its report on researchers' incentives and motivation for data sharing[30], and in actual fact there are research disciplines such as genomics and crystallography in which data sharing is so integrated that it has become a fully standardised research practice. The content of the interviews shows that to a lesser extent there is a tradition of data sharing within humanities and social science disciplines, while it is generally a more widespread aspect of the academic disciplines that fall under science and technical research. At DTU, for example, the perception is that the researchers generally have a positive opinion of the potential of Open Science.

[29] DTU on behalf of DEFF (2011): Research data and Open Access
[30] Knowledge Exchange (2014): Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective

> *"When we talk to the scientists they say that it is a promising idea, but it involves a lot of changes in culture and requires a strong commitment to use extra time on documentation of data."* (Focus group interview with the Research Data Group, DTU Library)

Several people also pointed out that the age of the researchers has an impact on their inclination to share data because many of the younger researchers are more accustomed to sharing information and data digitally and therefore also to pushing the boundaries.

> *"There is a great deal of interest among the younger researchers in particular. It's important for us to build upon this."* (Professor Peter Munk Christiansen, Head of the Department of Political Science at AU)

A tangible measure in the here and now that prompts the researchers to be more active with sharing and reuse of research data could be establishing more on-site support. As described above, Denmark needs to create and expand data supporter roles at the universities, while the evolution of DMPs can also set clear guidelines for the researchers in their work with research data. In addition to this, improved data access, including designing easy and user-friendly solutions for searching databases, can spur on the researchers to be more active with regard to sharing and reusing research data. Sharing knowledge and establishing an international context in the requirements set for data management means the evolution of DMPs can take place under the auspices of GO CHANGE, an initiative that addresses the culture in scientific research environments and the researchers' merit and incentive system.

3.2.3 Access

Access to digital research data is naturally a prerequisite for complying with the FAIR principles. Access to data can be qualified and, as emphasised by Barend Mons et al., there are degrees of FAIRness. FAIR is a continuum in which data must at the very least be Findable[31]. In addition to this, Accessible covers the idea that for each dataset it is possible to assess who is able to access data. This may be relevant in terms of medical research, for instance, where personal data need to be considered. On the other hand, it may be decided that data are accessible to all and therefore are fully open.

Experiences from Germany show that there is a need to clarify the principles more tangibly during the implementation and that in practice there are degrees of FAIRness. For example, those involved with RADAR have assessed the project against the FAIR principles and it is clear here that it is easier to comply with Findable and Accessible, whereas it has proven more difficult to achieve Interoperable and Reusable.

[31] Mons, Barend et al. (2017): Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Science Cloud

3.2.3.1 Requirements and considerations relating to technical solutions

A prerequisite for the researchers having access to research data is a research network that functions properly. Denmark has Forskningsnettet, a research network connecting the Danish universities and research institutions. Around 100 institutions are connected to the network, which is also linked internationally to NORDUnet and GÉANT. The connection to the international research networks is necessary to give the Danish researchers the opportunity to participate in international research collaborations. DeIC is responsible for Forskningsnettet in Denmark and its recommendations for a coherent e-infrastructure to promote Danish research contain an ambition for Danish research to become state-of-the-art in this field.

Furthermore, in relation to access, there is the development of the specific policies and procedures for data security, user authentication (that it is the correct user) and authorisation (what data a user can access). This is complicated further by the General Data Protection Regulation, which will come into force in May 2018.

> *"There may be different types of costs. People often forget the security costs, for example. Can data be altered by cyber crime? How do we manage personal data? If data are not open, how do we ensure that only a closed group of researchers can access them and not everyone?"* (Professor Børge Obel, Centre Director, DEFF)

Deliberations on access to data are a key part of the FAIR discussion because these are linked to data credibility and the quality of data. Therefore, it is in the first instance important for the technical solutions to support the right qualified researchers and subject experts with access to data, and in the second instance important to decide whether data have a quality and research-related interest that legitimises their storage and thus their reuse too.

> *"It must not always be easy to make data accessible. It should not be possible to simply save data if they are not approved for reuse. One of the major challenges is therefore to qualify the reuse."* (Lene Krøl Andersen, Head of eScience Competence Center, DeIC)

3.2.3.2 Research practice and division of responsibilities between researcher and institution

It is the individual researcher's research practice and management of data that determine whether other people are even given access to the data generated by that researcher. It was remarked in several interviews that one obstacle is the fact that researchers have not been made aware of the relevant legal rules on confidentiality, privacy or intellectual property rights (IPR) and that is why they often choose not to share their research data.

As a survey of research data management in Denmark has also demonstrated, the researchers often have sole responsibility for this decision because there are no institutional frameworks in place to support

them[32]. The researcher must thus personally ensure that research data are managed in line with best practice within the academic field, including making their own deliberations about the planning for the management of the primary materials and data, as well as on preservation of the research data, how long they will be stored, disposal of data, access to and sharing of data.

> *"There must be no doubt that data are correct. Who should validate the data that are accessible? All research is international. This is why we need to follow international technical standards, so we can share research data in a larger international network."* (Professor Børge Obel, Centre Director, DEFF).

The researchers can then be supported to a greater or lesser extent by the institutions' frameworks and policies; several Danish universities are gradually starting to focus on this. To take just one example, DTU has drawn up a data management policy in which it is responsible for providing a storage system that enables researchers to manage their data responsibly. This encompasses preservation, access control and back-up. The policy also sets out that DTU 1) must provide the opportunity for training and continuing training; 2) advise on practice, legal issues and infrastructures; and 3) develop and maintain common systems and infrastructures for research data management.

3.2.4 Infrastructure

The infrastructure is the foundation on which digital research data can be stored, managed and processed. Infrastructure consists of the local solutions at the universities, nationally in e.g. the Danish National Archives or Danish Data Archive and globally in international data repositories. In addition, infrastructure covers the necessary software that can e.g. support collaboration tools and advanced calculation programs for data processing, as well as calculation hardware used for modelling, simulation, data analysis and visualisation.

3.2.4.1 More investment needed in the establishment of more data storage and capacity

The interviews with the Danish universities draw a clear picture of the universities not possessing a complete overview of Danish storage capacity. There is not necessarily an overview internally at the universities either, because data storage and data management are dealt with locally at faculty and departmental level. Some departments have storage space they make available, while others do not. The recently completed project 'Data management in practice' has shown that there is a lack of data capacity in Denmark and there is certainly not a systematic approach to storing data.

> *"A few times we saw that they had stored data on USB hard drives. In other places our experience was that there was no control over it at all. While elsewhere we did have storage, but they didn't support the FAIR principles by any stretch of the imagination. So we helped them with this in the project*

[32] Thestrup B., Jesper et al. (2013): Management of research data in Denmark – A study of Danish universities' and research councils' practices, policy and visions for storage, long-term preservation, accessibility and sharing of research data

'Data management in practice'. A main conclusion from the project is that we need to find ways of making researchers more self-sufficient in relation to metadating." (Deputy Director General Bjarne Andersen, Head of IT Development and Infrastructure at the Royal Danish Library)

Even without a complete overview, a number of the experts interviewed believe that Denmark's data infrastructure is at present not sufficient to make FAIR a reality. The database re3data.org shows that Danish researchers are involved with 20 different data repositories both nationally and internationally. Of these 20, 14 do not use a PID (Persistent Identifier) system; the main problem associated with this is that it demonstrates that a very large proportion of the data storage used by Danish researchers is nowhere near meeting the criterion to be Findable and is therefore far from qualifying as FAIR data.

"The building blocks are in place in Denmark, albeit to varying degrees. To a great extent we have 'identifiers' in place that support findability, but we do not have the right infrastructure – that is where we are lagging behind." (Anders Sparre Conrad, Chairman of DM Forum).

Due to the lack of an overview, several people pointed out that the solution also ought to be found in the cooperation between universities. DM Forum is a platform that in this context should be used to ensure coordination across the educational establishments and assist this cooperation in producing viable solutions.

In relation to the GO FAIR initiative, which is based on GO CHANGE (dealing with the culture in scientific research environments and also the merit and incentive system for researchers), GO TRAIN (dealing with skills development) and GO BUILD (dealing with technical implementation), this indicates that Denmark should also focus on the GO BUILD element. On that point Germany is further ahead as it is the view of the stakeholders we interviewed that the infrastructure for storage and preservation of research data has been established in Germany.

3.2.4.2 Focus on long-term and sustainable funding for operation of data centres and preservation of data

DeIC recommendations for ensuring a coherent Danish e-infrastructure for the period 2018–2028 deem the current pilot projects at national level to be incapable of securing a sustainable infrastructure. It is therefore recommended that an amount matching the sums invested by our neighbouring countries be set aside to cover development, investment and operation of the current systems[33]. In other words, there is a need for more long-term funding for the development and operation of the infrastructure.

The focus in Germany on a long-term funding model represents a major challenge. A large proportion of the German investments are funded by project funds which do not cover the operating and maintenance costs. This is why it is the assessment of the German stakeholders that more long-term, institutional fund-

[33] DeIC (2017) Recommendations for a coherent e-infrastructure to promote Danish research

ing and/or thinking is needed in new business models, including the RADAR project, in which researchers and research institutions must pay to participate.

Generally speaking, the Danish stakeholders agree that there is a need to factor in long-term and sustainable funding for the operation of data centres and preservation of data etc. if FAIR is to be implemented successfully. In this context, some of the heads of department interviewed are concerned about a model in which researchers and research institutions are expected to pay in order to participate on too large a scale, as this will reduce the incentive to take part in the development of FAIR data. From this viewpoint, the interviewees highlighted the lack of relevant legislation. At the time of writing there are a number of publicly-funded national data archives, such as the Royal Danish Library, the Danish National Archives, etc., and several local university libraries working with research institutions that pay the costs of long-term preservation. In addition, the researcher is often personally responsible for the long-term preservation of the research results and for paying the data archive maintenance costs. The different ways of ensuring long-term preservation of data in use today show that the further concretisation of FAIR should concentrate on guaranteeing long-term and sustainable funding by recognising that long-term preservation of data is a resource-intensive service, the funding of which must be included in the total costs.

4. German experiences of introducing FAIR data

Overall, Germany is at the beginning of a process in which the FAIR principles need to be integrated and accepted by the key stakeholders, including universities, research libraries, data centres, etc. In 2017, Germany published a Joint Position Paper with the Netherlands calling for rapid action on the European Open Science Cloud (EOSC)[34]. In this paper Germany and the Netherlands voice the need for fast track implementation in order to expedite the development of EOSC by making data FAIR. The proposal is that this should happen by means of a decentralised approach in which the process is driven by the research communities themselves and builds on existing infrastructures and initiatives. At the same time, a main focus of the initiative is that national implementation should be coordinated and linked to the European implementation in order to avoid the duplication of efforts and investments.

The GO FAIR initiative (supported by the German Federal Ministry of Education and Research and headed by Leibniz Information Centre for Economics) is the focal point of the implementation of FAIR data in Germany. This initiative is based on three pillars: GO CHANGE (dealing with the culture in scientific research environments and also the merit and incentive system for researchers), GO TRAIN (dealing with skills development) and GO BUILD (dealing with technical implementation) respectively.

4.1 THE GERMAN STAKEHOLDER LANDSCAPE

The German research landscape contains a number of different types of R&D actors, including universities, non-university institutes, businesses and research institutions run by either federal or state authorities. Germany has approx. 1,000 publicly funded research institutions, approx. 400 of which are higher education institutions, including universities, university colleges and academies of professional higher education (under the umbrella name Fachhochschule (FHHS)). The German FHHS are important and integrated actors in the German research and innovation landscape.

[34] Joint Position Paper on the European Open Science Cloud (2017): Germany and the Netherlands

German research funding at federal level is handled by specific ministries, of which the Federal Ministry for Education and Research (BMBF) is the largest funding institution. The federal funds go to research at one of the around 750 research institutions in Germany or to thematic research within one of the key fields of research. In 2014 a total of approx. EUR 14 billion was awarded at federal level, with BMBF responsible for approx. 60 per cent of this figure. The German states are responsible for the remaining approx. 40 per cent of research funding and in 2013 handed out approx. EUR 10 billion.

A large proportion of research funding is being channelled through a series of large research funding organisations. Deutsche Forschungsgemeinschaft (DFG) is the largest independent research funding institution in Germany and in 2015 had a budget of approx. EUR 3 billion. Two thirds of DFG's funding comes from federal government and the final third from state government. The four major and internationally recognised research organizations Fraunhofer Gesellschaft, Max-Planck-Gesellschaft, Helmholtz Association and Leibniz Gemeinschaft also award funds through separate grant programmes.

4.2 A PARADIGM SHIFT IN THE GERMAN APPROACH

In Germany research data management really became part of the agenda in 2010 when the Alliance of Science Organisations in Germany (an association of the most important German research actors) drew up a set of ground rules for managing research data[35]. Since then other key actors in the German research system have focused on the issue, including the German Rectors' Conference (HRK, the association representing German universities in both the private and public sector); in 2014 this organisation drew up a series of recommendations for a more strategic approach to research data management[36]. Furthermore, a number of German states have launched several relevant initiatives: Helmholtz Association (made up of 18 different research centres) has supported projects in the domain, while DFG (the main German research funding institution) has developed guidelines as part of its research funding practice.

In Germany, the field of Open Science has been characterised by a large number of uncoordinated initiatives. It has largely been a bottom-up process but without any national coordination. There are many traditional and proud universities that are reluctant to cooperate and share their research data with other, rival universities. Within the states (Baden-Württemberg is as an example) it may be a challenge to get the universities to communicate.

The idea behind the GO FAIR initiative is to make a break with past practice in Germany; it is a policy vision to increase interdisciplinary coordination in order to avoid solutions that are mutually incompatible. Dr Klaus Tochtermann, Director of ZBW and in charge of the GO FAIR initiative, believes that the FAIR principles are a paradigm shift and a new way of managing research data. As he sees it, FAIR data are something different from Open Science.

> *"To meet the FAIR data and service vision, you need to have a global infrastructure in place, and build up the necessary skills among data stewards and researchers. One major difference in FAIR is the access part, as data does need to be intrinsically open."* (Klaus Tochtermann, Director of ZBW)

To demonstrate the value of FAIR data and get the principles embedded and operationalised, the German approach has been to launch projects that must demonstrate value and be best practice for what a model for the future German research data infrastructure might look like. In particular, the projects indicated below are key:

- **The Helmholtz Data Federation** (HDF), which is a consortium of six partners, consisting of high-performance computer centres and a community with a focus on big data which, due to their scope, are not transferable online.
- The **GeRDI** (Generic Research Data Infrastructure) project (funding from November 2016 to November 2019), whose long-term aim is to make it possible for all German researchers to store, share and reuse research data across fields of research. The first phase of the project will see the merger of

[35] Alliance of Science Organisations in Germany (2010): Grundsätze zum Umgang mit Forschungsdaten [Basic principles for management of research data]

[36] HRK (2014): Management of research data – a key strategic challenge for university management

three data centres. This should create the pilot case for roll-out in Germany being possible and serve as a model for the future German research data infrastructure.

- The **RADAR** (Research Data Repository) project (funding from 2013 to 2016) is all about collecting data in a data centre to which the different members have access. The project is interdisciplinary and its participants come from universities in different German states. The project is organised so that there is a local coordinator at each institution who is responsible for supporting the local researchers in data sharing and data management at their own institution. The project supports 'the long tail' within research data, which is those data that fall outside of the big data domain, i.e. first and foremost small datasets. In terms of FAIRness, the assessment is that the project has achieved the criteria of Findability and Accessibility but is far from meeting Interoperability and Reusability.

4.3 THE GERMAN FAIR DATA SOLUTION

The intention of the GO FAIR initiative is for it to be widely supported among the stakeholders and consensus-driven based on an acknowledgement that the implementation of FAIR is not just a matter of technical standards but is also largely conditional upon culture. As noted by the German informants, there is now a process underway whose aim is to 'convert' the German stakeholders, researchers and librarians so they understand and apply the FAIR principles. In this context it has been pointed out that the national policy focus and the national backing have played a decisive role in getting FAIR put on the agenda.

Unlike Denmark, Germany has a federal structure that gives the individual states policy-making capability to manage research policy; they are also able to launch initiatives such as eScience Initiative in Baden-Württemberg, HeFDI (Hessische Forschungsdateninfrastrukturen) in Hessen and Cooperation in Digital Science in Hamburg and Schleswig-Holstein themselves. Furthermore, the large research organizations have launched substantial investments in this area, including Helmholtz-Data-Federation at the Helmholtz Association, FORDATIS at Fraunhofer Gesellschaft and Leibniz Data at Leibniz Gemeinschaft. This increases the need for interdisciplinary coordination and dialogue.

4.3.1 The infrastructure is in place, but there is a need for more

The basic assessment of our informants is that the infrastructure for storage and preservation of research data has been established in Germany. There are data centres at 110 German universities, 233 technical educational institutions and 280 other research institutions, not to mention several thousand research libraries[37]. The next step is to build on this infrastructure, link them together and create consistent access for the researchers. Yet Dr Stefan Winkler-Nees from DFG is of the opinion that it will require a more large-scale, national coordination and strategy because right now, for example, the challenge is that the initiatives funded by the Helmholtz Association prioritise giving access to their research institutes.

There is furthermore a great need for investment in data supporters at universities in Germany. At present there is not a comprehensive overview of the situation in Germany, but all the stakeholders emphasised in their interviews that massive investment needs to take place in this area; at the same time, there is a challenge in getting universities and research libraries to increase their collaboration.

[37] Cf. the website re3data.org

"For research data management we need the libraries on board. In Germany, today about 20 per cent of the university libraries are already active, offering research data services to researchers and supporting the implementation of data management policies. Regional infrastructure programmes accelerate this rate (e.g. in Baden-Württemberg, North Rhine-Westphalia." (Birgit Schmidt, Scientific Manager, Göttingen State and University Library).

Delft University of Technology in the Netherlands is highlighted as best practice at a university that has invested in and worked systematically on providing data supporters for its researchers.

4.3.2 Need for clarification of the funding model

The funding of data centres and funding models in general are the subject of debate in Germany. There will obviously not be just a single funding model that is implemented, rather a combination of several different ones. Some data repositories will be funded by central government or state government, but a charge may be made to receive special services.

Another model is the membership model (tested in the RADAR project) in which several organisations join a consortium and each pays a membership fee to access data and the related services. Dr Stefan Winkler-Nees from DFG emphasises that this model is challenged by the fact that data centres find it difficult to price the costs and do not have an overview of the price of data.

A third model is the research foundations stipulating that the projects put their data in a data repository, with part of the research grant being earmarked for this. This is a practice followed by research funding institutions in the UK.

4.3.3 Technical standards are developed 'bottom-up'

As described, the German approach is to mature stakeholders and not stick to a large, all-in FAIR data solution that must then be rolled out across the whole country. The technical standards used must apply not only in Germany but also internationally. That is the reason why Germany (like Denmark) is taking part in initiatives such as the Research Data Alliance (RDA), whose purpose is to develop practices, policies and tools within a number of internationally appointed working groups and interest groups.

In its guidelines, DFG stipulates requirements for standards, PIDs and metadata forms but does not specify exactly what the applicants must use. This depends on tradition and practice within the specific research fields and sub-fields.

"There was some discussion about whether DFG should demand specific standards. However, we are convinced that it is up to the disciplines themselves what type of standards they will use. Standards can be very complex and specific, and they may need a long time to develop. But it is important that there is progress, driven by a bottom-up approach." (Stefan Winkler-Nees, Programme Director at DFG)

4.4 WHAT CAN DENMARK LEARN FROM THE GERMAN EXPERIENCES?

In our opinion, Denmark has the opportunity to build on the German experiences of implementing a FAIR data solution:

Firstly, it is absolutely essential to maintain focus on demand from the researchers as a driver for the initiative. In practical terms, this means that a FAIR data solution must be developed with an appreciation for the researchers' use and needs, e.g. whether they are adequately supported in the changeover to working with FAIR data and whether the solutions developed consider their practices. In Germany there is focus on dialogue with the research environments concerning the potential of FAIR and Open Science; to mention just one example, the GO FAIR initiative focuses on 'maturing' the research environments and engaging in dialogue with the researchers about the potential of FAIR.

Secondly, it is a challenge in Germany to get assured sustainable funding of FAIR. Many of the projects awarded grants have been funded for a limited time period and the risk inherent in project financing is that the projects will founder as soon as the money runs out. In Germany, discussions are ongoing on which funding models people expect to be implemented to ensure long-term and sustainable funding. We believe it to be important for Denmark to also start a discussion on funding models, including how they can contribute to long-term, sustainable funding for infrastructure, training, data support, etc.

Thirdly, experiences in Germany show that the path from making a decision to become FAIR to realising the vision is a long one. The German research system is more fragmented than the Danish system and consists of around 1,000 publicly funded research institutions. At the same time, the German system is federal, meaning that the individual states have policy decision-making authority for research priorities. It must also be expected that it will take a long time to achieve the vision of FAIR data in Denmark, even though it has a smaller policy structure and research system, which is why a solution can be rolled out more quickly.

5. Methodology and data collection

The analysis is based on a number of different methodologies. The estimation of costs and benefits from the introduction of FAIR data in Denmark is primarily based on a thorough literature study of benefits and costs associated with the introduction of FAIR data, but we have taken a critical view of the literature in terms of providing perspective in a Danish context. The collection of German experiences and assessment of the Danish opportunities and barriers are based on a combination of a thorough desk research of studies, reports and articles, as well as a wide-ranging interview study focusing on key stakeholders' experiences and knowledge of FAIR and their view of the potential for introducing FAIR.

The methodological approach to each of the interdisciplinary analyses is described in depth below with regard to estimating the cost and benefits (section 5.1) and also the opportunities and barriers in introducing FAIR data (section 5.2).

5.1 ESTIMATION OF COSTS AND BENEFITS

This section provides an account of the different methods that can be used to assess the value of introducing FAIR data in Denmark and of the selected method in the socio-economic impact assessment in Chapter 2.

The methodology is based on a thorough literature review. This literature review has mainly focused on methodologies that have previously been used to assess the value of FAIR data but it also includes literature about the value of other types of open data (e.g. open public sector data). The methodologies from this literature have been brought in where they are relevant to FAIR data. Finally, where relevant, we have incorporated knowledge from interviews in Denmark and Germany to refine our methodology.

5.1.1 General approaches to socio-economic impact assessments

Our overall approach to socio-economic impact assessments follows the guidelines issued by the Danish Ministry of Finance, which is the standard in Denmark. There are very few examples of socio-economic impact assessments of FAIR data in the literature and no examples that include start-up or operating costs. This means that we found it necessary to draw on a number of different sources and methodologies in our work on estimating the socio-economic value of introducing FAIR data in Denmark.

This applies to e.g. the timing of benefits and costs respectively. The return on research and thus the benefits of FAIR data are modelled over a 10-year period in Houghton & Gruen (2014). In several other studies in the literature a 30-year period is used. We use the same time horizon for the return on research as in Houghton & Gruen (2014). The literature on the timing of the costs, including start-up and operating costs, is very limited. That is why we have decided to make use of fairly conservative assumptions.

Another challenge is that a large part of the literature in this area consists of ex post analyses or evaluations of existing FAIR data solutions. A considerable limitation of the existing ex post studies is that they measure the value of data when they *are* FAIR and not – which is the purpose of this analysis – the added value of *introducing* FAIR data. There are, however, individual examples of ex ante analyses that work out the value of introducing new FAIR data schemes or the roll-out of existing schemes. In ex ante studies, the value of data is estimated when they are not FAIR; different methodologies are subsequently used to assess the potential added value that data will acquire when they become FAIR. Our overall approach is

therefore largely based on these ex ante studies. The Houghton & Gruen (2014) study in particular is the main source for the estimation of benefits from FAIR data.

5.1.2 Estimation of benefits from FAIR data

In the literature there are basically two different methodologies that can be used to estimate the value of data: the welfare theoretical methodology and the return-based methodology.

The welfare theoretical methodology estimates consumer surplus (the user's willingness-to-pay minus the actual payment) on the basis of price elasticities. This approach is not, however, particularly suitable when the price of data is reduced to 0, i.e. data are free of charge. Furthermore, the data necessary to estimate price elasticities are not fully accessible for research data in Denmark.

In contrast, the starting point for the return-based methodology is the literature on the return on research and development. Good data on the research activity in Denmark and an existing literature on the return on research and development in Denmark mean that this methodology is more suitable for working out the value of introducing FAIR data in Denmark. The majority of the existing literature is also based on this approach.

Figure 5.1. Estimation of the return on publicly funded research

The return is estimated according to the same method used in Houghton & Gruen (2014) and is obtained for the total supply of research data, which increases every year once new data are generated. Data **are discounted by 10 per cent per annum**, based on literature indicating a rate of depreciation of 5–10 per cent (Evans et al., 2008). Our assumption here is that data have a **useful life of 10 years**, although several studies do identify a useful life of up to 30 years (Sveikauskas, 2007). The return is expected to be realised over five years and has a normal distribution over years 1 to 5 (Mansfield, 1991; Mansfield, 1998).

As mentioned, we use Houghton & Gruen (2014) as our starting point in our estimate of the benefits of FAIR data. This is the only study produced ex ante while at the same time being based on the return-based methodology. We believe that applying this methodology gives us the best opportunity to allow for the fact that a large amount of data already exists in Denmark, and that the value of this should not be included when working out the socio-economic value of introducing FAIR data in Denmark.

The starting point for estimating the benefits is the value of research data in Denmark. In addition to this, there is a host of secondary benefits from data becoming FAIR, including the value of time saved and the creation of new research. According to our interviews, these are the benefits for which there is the greatest potential in Denmark. A possible benefit of FAIR data is that the return on research may be higher because research is made more targeted or interdisciplinary. This benefit is not included in our main analysis because we want to make a conservative estimate of the benefits and because the benefit does not appear to be significant in our interviews with actors in Germany and Denmark.

The literature contains studies of FAIR data solutions for different branches of science. FAIR data are clearly most widespread for natural science, and consequently there is also much more literature in this area. Yet studies do indicate that there is a difference in the effects for natural science, social science and the humanities. Our calculations of the benefits are therefore based on a division into these three disciplines.

Since there are no separate studies for technical sciences, health sciences and agricultural and veterinary sciences, these are calculated on the basis of the same estimates as for natural science.

5.1.2.1 The value of research data

Being able to estimate the benefits of making data FAIR requires the value of the existing research data in Denmark to be estimated first. The literature contains a number of different methodologies for estimating the value of research data. Several of these are based on data which cannot be accessed by Denmark. The most important ways of assessing the value of research data are set out briefly below:

Investment value: This expresses the total value of the investment in FAIR data, including operation of infrastructure, guidance and service. It does not contain sunk costs, i.e. one-off investments for setting up infrastructure etc. It is based on the literature on operating accounts from data centres, as well as questionnaires about costs associated with uploading data to existing data centres. This methodology is best suited to ex post analyses.[38]

Willingness-to-pay value: Based on stated preferences, data are valued according to the users' willingness to pay. 'Stated preferences' is a methodology in which users of data are asked to evaluate the value of data in a questionnaire.[39]

Use value: The value of data is estimated as the costs associated with the use of data, including in particular the value of time spent on data work but also costs for e.g. data analysis tools etc. In ex post analyses, this is estimated on the basis of questionnaire surveys given to users from the data centres being analysed. Houghton & Gruen (2014) estimate the utility value based on macroeconomic data about the national research and development expenditure, as well as questionnaire surveys from existing data centres on the proportion of research time that is spent on data work. Macroeconomic data for research and development expenditure are available for Denmark and are of high quality; that is why we use this methodology in our analysis.[40]

Houghton & Gruen (2014) work with two estimates for the use value of data. The first is the upper bound estimate, based on total expenditure on research and development. By including the total research and development expenditure, costs associated with data analysis tools or other capital costs associated with the use of data are also incorporated. The estimate may, however, be an upper bound estimate, as there is no empirical evidence for what proportion of the capital costs for research and development is related to data work. In contrast, this evidence exists for the wage costs or time costs, based on a number of different questionnaire surveys for existing data centres. They consequently use the proportion of the wage costs related to data as the lower bound estimate. In our analysis, the starting point is the wage costs, corresponding to the lower bound estimate in Houghton & Gruen (2014).

[38] Used in Sullivan et al. (2017), Houghton & Gruen (2014), Beagrie & Houghton (2014), Beagrie & Victoria University Centre for Strategic Economic Studies (2012)
[39] Used in Sullivan et al. (2017), Beagrie & Houghton (2016), Beagrie & Houghton (2014), Beagrie & Victoria University Centre for Strategic Economic Studies (2012)
[40] Used in Sullivan et al. (2017), Beagrie & Houghton (2016), Beagrie & Gruen (2014), Beagrie & Houghton (2014), Beagrie & Victoria University Centre for Strategic Economic Studies (2012)

5.1.2.2 The value of time saved

The introduction of FAIR data is expected to result in time savings because researchers will spend less time on acquiring and structuring data. This in itself has a value, as researchers' time is valuable and can be given over to other activities. Quicker access to data makes the research and development process faster and increases the return on research, for instance by researchers 'investing' the time saved in other research that then produces an additional return.

The approach to estimating the time saving associated with FAIR data in Houghton & Gruen (2014) is based on experiences from the UK, where various FAIR data solutions have been implemented. We apply the same methodology but use multiple sources for the time saving estimate.

In some studies only the value of the time saving (measured as a DKK saving in wages) is included as an efficiency gain. In Houghton & Gruen (2014), the additional return on the extra research that the researchers invest their time in is included too. We use the same approach in our analysis.

In some studies the efficiency gains are considered more broadly, i.e. not exclusively for researchers but also for teachers and students who also use these data.[41] These efficiency gains are not included in our analysis.

5.1.2.3 The value of new research produced as a result of FAIR data

Wider access to data offers better opportunities for interdisciplinary research and cooperation across institutions and sectors. New research that not would arise without FAIR data may then emerge. The better access may also facilitate researchers with smaller budgets making use of data, given that they save time and costs associated with producing data. The interviews indicate that the potential for interdisciplinary research may be limited, but there is still thought to be major potential for creating new research as a result of FAIR data in Denmark.

Houghton & Gruen (2014) estimate the amount of new research in the light of the questionnaire surveys for existing projects in the UK. These indicate what proportion of the research would not be completed without FAIR data. In Houghton & Gruen (2014), the full value of this new research is recognised as a benefit of FAIR data. This is equivalent to assuming that the research activity increases substantially because the assumption is that the research which will take place when data are not FAIR continues to be carried out, at the same time as new research taking place on the basis of FAIR. In other words, the assumption is that new research does not displace existing research. We consider this to be quite an optimistic assumption. Since there is no literature that empirically supports the degree of displacement of existing data, and since we want to follow a conservative approach, in our main scenario we have therefore chosen to disregard the value of new research produced as a result of FAIR data. Sensitivity analyses on the importance of new research were performed though.

5.1.2.4 Greater return on research

Easier access to data may result in better and more informed research, including a reduction in the time invested in blind alleys and the duplication of research. This can help to avoid resources being wasted and costs for the same research being doubled. Research will be more efficient and thereby increase the aver-

[41] See Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)

age return on the research. Taken as a whole, FAIR data may lead to more efficient research through a number of channels: better teaching, research rolled out more widely, greater interdisciplinary collaboration, etc.

This benefit is described in the literature, but it is difficult to estimate its precise scope and therefore the value. A macroeconomic approach makes it possible to illustrate the potential of more efficient research. A calculation of the potential is made in Houghton & Sheehan (2009) which is based on a modified Solow model. The study shows how accessibility and efficiency (proportion of new knowledge that is useful) are important parameters when estimating the return on research. Thereby, they show theoretically that an increase in accessibility and efficiency increases the return on research. Yet since the study does not contain suggestions for proxies for the parameters of accessibility and efficiency, it is not possible to assess empirically to what extent these parameters are influenced by the introduction of FAIR data. The benefit is not included in our analysis due to these uncertainties.

5.1.3 Estimating the costs of FAIR data

The literature has less focus on estimating the costs of FAIR data. This may be related to the fact that the vast majority of the studies are ex post studies; therefore, in these studies the costs can easily be calculated using the financial accounts for the individual projects and one can disregard the start-up and investment costs that may be deemed sunk costs[42]. Houghton & Gruen (2014) estimate the operating costs based on their share of the value of research data. The study does not, however, include start-up or investment costs and does not model the costs over time, which it is necessary to do in our analysis.

Beagrie et al. (2010) map all the potential types of costs related to the preservation of research data on the basis of several different projects and data centres. The study focuses on classifying and analysing the types of costs. In most cases, the total data value for the data centres is not stated, making it difficult to extrapolate the costs for Denmark. Having said that, the study does feature one case that analyses the distribution of the total data costs. In connection with the Embedding Institutional Data Curation Services in Research (EIDCSR) project at Oxford University, information has been collected about all costs associated with data work, including the costs of generating data and both start-up and operating costs for the data centre. We use this distribution to estimate the start-up and operational costs as a proportion of the total data value based on the same approach used in Houghton & Gruen (2014).

Our estimates thus take into account that the introduction of FAIR research data in Denmark will involve high start-up costs. Estimates are conservative in the sense that costs will likely be overestimated. This is due to the study only covering the costs associated with the creation of data, while the total data value in our analysis includes all costs associated with the use of data, including the costs of analysing data. This means the proportion of start-up and operating costs in relation to the total costs will be overestimated. What is more, these proportions are based on a shorter project period than our analysis period. It may therefore be the case that we include start-up costs for too long a period and/or that the start-up costs will fall over time. The proportions are used based on a conservative approach.

[42] 'Sunk costs' is a concept used by economists to describe costs that have already been realised and therefore should not influence decision-making.

5.1.4 Alternatives to cost-benefit analysis

The literature contains numerous alternative studies on the value of FAIR research data, but these cannot be used directly in a cost-benefit analysis. The various methodologies and approaches used in the literature are summarised briefly here.

5.1.4.1 Extrapolation based on GDP

In the European Commission study (2015), the value of data is extrapolated based on macroeconomic figures for the value of data as a percentage of GDP, with the maturity level of open data in the different EU countries also being taken into account. The study looks at both the direct and indirect value of open data. In relation to our analysis, a challenge of this approach is that it is difficult to distinguish between FAIR research data and other types of open data. The study is based on an ex post approach. Given that we want to isolate the socio-economic value of introducing the FAIR principles in Denmark, it is also important to correct for the value of the existing data. This is not done in the study.

A similar method is applied more widely in the OECD study (2015) for public data (PSI) and in Gruen (2014) for open data.

5.1.4.2 Estimates based on cases

Gruen (2014) refers to a variety of different cases, with examples of open data, to estimate the value of open data. Estimates are based on a number of very uncertain assumptions and should therefore be regarded as calculations of the potential rather than genuine socio-economic impact assessments. Gruen (2014) mainly uses the methodology to argue that the result, which is based on extrapolation from GDP, is not unrealistically high.

5.1.4.3 Return on investment (RoI) from analyses of the value of libraries

It is emphasised in Beagrie & Victoria University Centre for Strategic Economic Studies (2012) that the literature on the return on investments in data, library and information services is relatively well-established. This literature can serve as an additional check of our calculations, if the return on investments in FAIR data is expected to be around the same size. The literature finds that investments in data, library and information services create a value between approx. three and five times the size of the investment.[43]

[43] Source: Beagrie (2012)

5.1.5 Sources used in our calculations

Parameters and inputs	Source(s)
Discount rate	Danish Ministry of Finance (2017)
Tax distortion	Danish Ministry of Finance (2017)
R&D expenditure as a percentage of GDP	Statistics Denmark, statistikbanken.dk
R&D FTEs (faculty members) in the public sector by gender, category of staff, field of science and time	Statistics Denmark, statistikbanken.dk
Standard calculated monthly earnings, public administration and service	Statistics Denmark, statistikbanken.dk
GDP in current prices	Statistics Denmark, statistikbanken.dk
Real growth in GDP	Danish Ministry of Finance (2017) – updated 2025 projection
Proportion of researchers' time spent on research and knowledge creation	DEA (2015)
Proportion of research time spent on data work	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Efficiency gain, proportion of time spent on data work that can be saved	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b), Sullivan et al. (2017)
Return on publicly funded research	DEA (2017), Universities Denmark (2016), Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b), Sullivan et al. (2017)
Proportion of return that can be achieved locally	Houghton & Gruen (2014), Jaffe (1989), Coe & Helpman (1993, 1995), Verspagen (2004), Arundel & Guena (2004)
Rate of depreciation for research	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Useful life of data	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Proportion of research that would or could not be produced without FAIR data	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Victoria University Centre for Strategic Economic Studies (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Operating costs	Houghton & Gruen (2014), Beagrie et al. (2010)
Start-up costs	Beagrie et al. (2010)

5.2 ANALYSIS OF OPPORTUNITIES AND BARRIERS

This section provides an account of the selected methodologies and the data collected in the interdisciplinary analysis of barriers and opportunities (Chapter 3), as well as of the German experiences of introducing FAIR data (Chapter 4).

5.2.1 The German case

Germany has been selected as a case study in the preliminary analysis, the reason being that Germany is one of the 'early movers' countries that has set about carrying out the GO FAIR initiative and has decided to implement the FAIR principles on the basis of the EOSC High Level Expert Group's recommendations.

We have used the case study as a way of capturing the German experiences in depth from different angles and perspectives. Collecting knowledge about the German experiences has been a focal point of the first phase of the preliminary analysis. The background to the mapping of the most important German experiences is firstly that it has allowed us to analyse the opportunities for transferring the German experiences to a Danish context and assessing whether benefits, costs, opportunities and barriers will also apply in Denmark.

The case study was compiled through a thorough and systematic desk research of relevant reports, articles, etc., in addition to six in-depth, qualitative interviews with key German stakeholders. The desk research phase consisted of reports, strategies and articles relating to the area, including:

- German Federal Ministry for Education and Research (2016): Open Access in Deutschland – Die Strategie des Bundesministeriums für Bildung und Forschung [Open Access in Germany – The Strategy of the Federal Ministry for Education and Research]
- German Council for Scientific Information Infrastructures (2016): Enhancing Research Data Management: Performance Through Diversity
- Alliance of Science Organisations in Germany (2015): Position paper 'Research data at your finger tips'
- Germany and the Netherlands (2017): Joint Position Paper on the European Science Cloud
- German Rectors' Conference (2014): Management of research data – a key strategic challenge for university management

Based on the above reports relating to the area and the Danish Agency for Science and Higher Education's contact with the relevant organisations involved in the German effort, six interviewees were selected, as shown below:

Table 5.1 Overview of interviewees in Germany

Organisation	Interviewee	In person/by telephone
Deutsche Forschungsgemeinschaft (DFG)	Dr Stefan Winkler-Nees, Director	Two interviews conducted (in person and by telephone)
FIZ Karlsruhe	Matthias Razum	In person
ZBW (Leibniz Information Centre for Economics)	Prof Dr Klaus Tochtermann	In person
Leibniz Universität Hannover	Prof Dr Eng. Gabriele von Voigt	By telephone

Göttingen State and University Library	Dr Birgit Schmidt	By telephone

We did not manage to conduct interviews with some of the German actors, e.g. the German Federal Ministry of Education and Research. We have tried several times to contact them by telephone and email but we were unsuccessful. Nevertheless, we have the identified the Ministry's approach to FAIR data via the other interviews and a thorough desk research of relevant strategy papers and reports from the Ministry.

The interviews were conducted following a semi-structured interview guide adapted to the individual informant's role and knowledge of Germany's efforts. They were conducted in person in Germany in order to capture the experience of key actors in depth.

5.2.2 Danish stakeholders' attitude to FAIR data

The starting point for the interdisciplinary analysis of barriers and opportunities in respect of FAIR data in Denmark is a thorough desk research of relevant reports, analyses and other related literature, including:

- DeIC, DEFF (2015), *National strategy for research data management 2015–2018*, Copenhagen, 30 January 2015
- Horizon 2020 Programme (2016), *Guidelines on FAIR Data Management in Horizon 2020*
- Force11 (2017), *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing*
- European Commission (2017), *European Open Science Cloud*
- Max Planck-Gesellschaft (2002), *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*
- Mons, Barend; Neylon, Cameron; Velterop, Jan; Dumontier, Michel; da Silva Santos, Luiz Olavo Bonino & Wilkinson, Mark D. (2017), *Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud*
- OECD (2007), *Principles and Guidelines for Access to Research Data from Public Funding*
- Danish Government (2016), *Science, Technology and Innovation Policy report*
- Wessels, Bidgette; Finn, Rachel L.; Wadhwa, Kush; Sceinsdottir, Thordis; Bigagli, Lorenzo; Nativi, Stefano & Noorman, Merel (2017), *Open Data and Knowledge Society*, Amsterdam University Press
- Wilkinson et al. (2016), *The Fair Guiding Principles for Scientific Data Management and Stewardship*.

This desk research into the area initially in the preliminary analysis led to a background memo that has described different definitions of FAIR data and also the background to the desire to introduce FAIR data at European level. Parts of this background memo have been incorporated into this report.

A great many interviews have been conducted with relevant Danish stakeholders as a key part of the analysis of barriers and opportunities in implementing FAIR data in a Danish context. We carried out a total of 30 in-depth, qualitative interviews with actors at policy level (including the EU) and system level (including coordinating actors and national research libraries), as well as with research funding institutions, management and administrative staff from the Danish universities, not to mention researchers. In our analysis we made deliberate attempts to come into contact with a wide spectrum of people from the Danish universities, among other things for the purpose of creating ownership and backing for the analysis and to ensure that everyone feels involved.

Table 5.2 Overview of interviewees in Denmark

Organisation	Department	Interviewee	Title	Type
EOSC	High Level Expert Group	Barend Mons	Chairman	EU
Danish Agency for Digitisation	Division for Data and Architecture	Jens Krieger Røyen	Head of Division for Data and Architecture	Policy level
Danish Health Data Authority	Board of Directors	Lisbeth Nielsen	Director General	Policy level
Denmark's Electronic Research Library (DEFF)		Børge Obel	Centre Director	System level
Royal Danish Library	IT Development and Infrastructure	Bjarne Andersen	Deputy Director General, Head of IT Development and Infrastructure	System level
Coordinating Body for Register Research (KOR)		Jeppe Klok Due	Special adviser	System level
National Forum for Research Data Management (DM Forum)		Anders Sparre Conrad	Chairman	System level
Research Data Alliance (RDA)/DeIC		Lene Krøl Andersen	Head of eScience Competence Center	System level
Danish e Infrastructure Cooperation (DeIC)		Steen Pedersen	CEO	System level
Danish National Archives	Senior management	Kirsten Villadsen	Deputy Director	System level
Danish Council for Research and Innovation Policy (DFiR)		Jens Oddershede	Chairman	Research funding institution
Innovation Fund Denmark	Management	Peter Høngaard Andersen	Director	Research funding institution
Novo Nordisk Foundation		Thomas Alslev Christensen	Head of Operations	Research funding institution
Copenhagen Business School (CBS)	Research Support Office	Lars Nondal	Chief Adviser	Research institution
Technical University of Denmark (DTU)	Office for Innovation and Sector Services	Katrine Flindt Holmestrand	Project manager	Research institution
Technical University of Denmark (DTU)	Office for Innovation and Sector Services	Paula Maria Martinez Lavanchy	Project team member	Research institution
Technical University of Denmark (DTU)	Office for Innovation and Sector Services	Falco Jonas Hüser	Project team member	Research institution
Universities Denmark	Danish Rectors' Conference	Nikolaj Helm-Petersen	Deputy Head	Research institution
The Royal Danish Academy	Research and artis-	Henrik Oxvig	Head of Research	Research institu-

of Fine Arts Schools of Architecture, Design and Conservation (KADK)	tic development services			tion
University of Copenhagen (UCPH)	Science	John Renner	Dean	Research institution
Roskilde University (RUC)	Administration	Peter Kjær	Pro-rector	Research institution
IT University of Copenhagen	Research administration	Jens Christian Godskesen	Pro-rector	Research institution
Statens Serum Institut (SSI)		Mads Melbye	Sector Director	Research institution
University of Southern Denmark (SDU)	University library	Bertil F. Dorch	Library Director	Research institution
Aalborg University (AAU)	Technical Faculty of IT and Design	Henrik Pedersen	Dean	Research institution
Aarhus University (AU)	Aarhus University, School of Business and Social Sciences	Peter Munk Christiansen	Head of Department	Research institution
University of Copenhagen (UCPH)	Department of Public Health	Merete Osler	Professor	Researcher
University of Copenhagen (UCPH)	Humanities – Department of Nordic Studies and Linguistics	Bente Maegaard	Senior Executive Adviser	Researcher
University of Copenhagen (UCPH)	Saxo Institute, incl. Centre for Textile Research	Christina Margariti	PhD	Researcher
University of Southern Denmark (SDU)	Department of Public Health	Kirsten Ohm Kyvik	Head of Department	Researcher

The interviews were mainly conducted over the telephone following a semi-structured interview guide adapted to the individual actor type. The primary purpose of the interviews was to obtain details of the informants' assessment and use of open research data in general, including how they work with open research data. Questions were then asked about the informants' assessment of the FAIR principles as well as the potential for introducing them in Denmark, including possible benefits and costs/challenges. In each interview the consultant outlined the German case for implementing FAIR data with a focus on 1) the model for FAIR data; 2) opportunities/barriers and; 3) benefits. The informant was subsequently asked to assess whether and how the German experiences can be transferred to a Danish context.

6. References

Mansfield, E. (1991), 'Academic research and industrial innovation', Research Policy 20(1), 1991, pp. 1–12.

Mansfield, E. (1998), 'Academic research and industrial innovation: an update of empirical findings', Research Policy 26(7/8), 1998, pp. 773–776.

Evans, P., Hatcher, M. and Whittard, D. (2008), The preliminary satellite account for the UK: a sensitivity analysis, Economic & Labour Market Review 2(9), September 2008, pp. 37–43.

Sveikauskas, L. (2007), R&D and Productivity Growth: A Review of the Literature, US Bureau of Labor Statistics, Washington DC, Working Paper 408.

Danish Ministry of Finance (2017), Vejledning i samfundsøkonomiske konsekvensvurderinger [Guidance on socio-economic impact assessments].

Sullivan, K. P., Brennan-Tonette P. & Marxen L. J. (2017), Economic Impacts of the Research Collaboratory for Structural Bioinformatics.

Beagrie N. & Houghton J. (2016), The Value and Impact of the European Bioinformatics Institute.

European Commission (2015), Creating Value through Open Data.

OECD (2015), Assessing government initiatives on public sector information.

Houghton, J. W. & Gruen, N. (2014), Open Research Data.

Beagrie N. & Houghton J. (2014), The Value and Impact of Data Sharing and Curation. A synthesis of three recent studies of UK research data centres.

Lateral Economics (2014), Open for Business: How Open Data Can Help Achieve the G20 Growth Target.

Schrier, B. (2014), Government Open Data: Benefits, Strategies and Use.

Beagrie, N. and Houghton J. W. (2013a), The Value and Impact of the Archaeology Data Service: a study and methods for enhancing sustainability.

Beagrie, N. and Houghton J.W. (2013b), The Value and Impact of the British Atmospheric Data Centre.

McKinsey (2013), Open Data: Unlocking innovation and performance with liquid information.

De Vries (2012), Re-use of Public Sector Information – Catalogue and highlights of studies, cases and key figures on economic effects of changing policies.

Beagrie, N. & Victoria University Centre for Strategic Economic Studies (2012), Economic Impact Evaluation of the Economic and Social Data Service.

Houghton, J. W., Swan, A., & Brown, S. (2011), Access to Research and Technical Information in Denmark, the Danish Agency for Science, Technology and Innovation and Denmark's Electronic Research Library.

Houghton, J. W. (2011), Costs and Benefits of Data Provision.

Vickery (2011), Review of recent studies on PSI re-use and related market developments.

Zangenberg and Co. (2011), Kvantificering af værdien af åbne offentlige data [Quantification of the value of open public sector data].

Beagrie, N., Lavoie, B. & Woollard, M. (2010), Keeping Research Data Safe 2.

Houghton, J. W. & Sheehan P. (2009), Estimating the Potential Impacts of Open Access to Research Findings.

Beagrie, N. (2009), Draft Guide to Cost/Benefit Analyses for Research Data Services.

Fry, J., Lockyer, S., Oppenheim, C., Houghton, J. W., & Rasmussen, B. (2008), Identifying Benefits Arising from the Curation and Open Sharing of Research Data Produced by UK Higher Education and Research Institutes, London and Bristol: JISC.

Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006), MEPSIR: Measuring European Public Sector Information Resources, Brussels: European Commission.

PIRA (2000), Commercial Exploitation of Europe's Public Sector Information.

Jaffe, A. (1989), 'Real effects of academic research', American Economic Review 79, pp. 957–970.

Coe, D. T. and Helpman, E. (1993), International R&D Spillovers, NBER Working Paper 4444, National Bureau of Economic Research, Cambridge MA.

Coe, D. T. and Helpman, E. (1995), 'International R&D Spillovers', European Economic Review 39, pp. 859–887.

Arundel, A. and Geuna, A. (2004), 'Proximity and the use of public science by innovative European firms', Economics of Innovation and New Technology 3(6), pp. 559–580.

Verspagen, B. (2004), The Impacts of Academic Knowledge on Macroeconomic Productivity Growth: An Exploratory Study, Eindhoven Centre for Innovation Studies, Eindhoven.

DeIC, DEFF (2015), National strategy for research data management 2015–2018, Copenhagen, 30 January 2015

Horizon 2020 Programme (2016), Guidelines on FAIR Data Management in Horizon 2020

Force11 (2017), Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing

European Commission (2017), European Open Science Cloud

Max Planck-Gesellschaft (2002), Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

Mons, Barend; Neylon, Cameron; Velterop, Jan; Dumontier, Michel; da Silva Santos, Luiz Olavo Bonino & Wilkinson, Mark D. (2017), Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud

OECD (2007), Principles and Guidelines for Access to Research Data from Public Funding

Danish Government (2016), Science, Technology and Innovation Policy report

Wessels, Bidgette; Finn, Rachel L.; Wadhwa, Kush; Sceinsdottir, Thordis; Bigagli, Lorenzo; Nativi, Stefano & Noorman, Merel (2017), Open Data and Knowledge Society, Amsterdam University Press

Wilkinson et al. (2016), The Fair Guiding Principles for Scientific Data Management and Stewardship.

Pryor, Graham (2012): Why manage research data?

German Federal Ministry for Education and Research (2016): Open Access in Deutschland – Die Strategie des Bundesministeriums für Bildung und Forschung [Open Access in Germany – The Strategy of the Federal Ministry for Education and Research]

German Council for Scientific Information Infrastructures (2016): Enhancing Research Data Management: Performance Through Diversity

Alliance of Science Organisations in Germany (2015): Position paper 'Research data at your finger tips'

Germany and the Netherlands (2017): Joint Position Paper on the European Science Cloud

German Rectors' Conference (2014): Management of research data – a key strategic challenge for university management

OXFORD RESEARCH

DENMARK
Oxford Research A/S
Falkoner Allé 20
DK-2000 Frederiksberg
Denmark
Tel: (+45) 3369 1369
office@oxfordresearch.dk

NORWAY
Oxford Research AS
Østre Strandgate 1
NO-4610 Kristiansand
Norway
Tel: (+47) 4000 5793
post@oxford.no

SWEDEN
Oxford Research AB
Norrlandsgatan 11
SE-103 93 Stockholm
Sweden
Tel: (+46) 08 240 700
office@oxfordresearch.se

FINLAND
Oxford Research Oy

Fredrikinkatu 61a, 6krs.
00100 Helsinki, Finland
www.oxfordresearch.fi
office@oxfordresearch.fi

BRUSSELS
Oxford Research
C/O ENSR
5. Rue Archiméde
PO Box 4, 1000 Brussels
www.oxfordresearch.eu
office@oxfordresearch.eu