Forskningsrådet

Investment in infrastructures for FAIR research data and public administration data of particular relevance to research

The organisation and funding of data infrastructure to ensure best possible use

Recommendations from the Data Infrastructure Committee
May 2022

Contents

Preface

Data has always been central to research but has often been the resource and responsibility of individual researchers. Technological developments have led to far more and increasingly advanced sources of data, often from large infrastructure platforms with many users. This in turn increases the cost of collecting data. The rapid pace of digitalisation has provided new opportunities for sharing data, which has also led to an increased recognition of the value of research data. This highlighted the need for publicly funded research data and public administration data of particular relevance to research to be shared to allow the public investment to benefit all, provide good opportunities for verification of research, and to let society reap the inherent benefits of publicly funded research and public administration data.

This work must be seen in the context of other reports that have proposed solutions in various areas to increase sharing and reuse of data. There is a great deal of overlap in recommendations and together they form a good basis for pursuing action, so that the value of the data is better utilised in the future.

Norway is well positioned to lead the way in this work, with the benefits this will bring to Norwegian research, the public administration and business development, but it will require political will and implementation capacity across sectors, as well as good cooperation between various actors.

I would like to thank the members of the Committee, who have contributed their expertise and experience from various professional and administrative areas. Thank you also to the external resource persons who provided well-prepared and pertinent contributions during the process. The remit has been extensive and, with a relatively short time horizon coupled with restrictions on physical meetings, discussion has focused more on general principles. This has resulted in a proposed framework for the organisation and funding of data infrastructures in Norway.

On behalf of the Committee, I would like to thank the Research Council of Norway and the Ministry of Education and Research for this interesting assignment. I would also like to extend a big thank you to the secretariat, Rita Bergersen, Siri Lader Bruhn and Katrine Weisteen Bjerde, for their invaluable support to the Committee's work. Finally, thank you to the many people who have shared their knowledge and expertise with the Committee during our work.

Kenneth Ruud
Chair of the Committee

10 May 2022

Summary

In order to achieve high-quality research, it is important that the data used are quality assured and can be used by others. That is how we generate research that can be trusted, and in which people have confidence. Unfortunately, there are too many examples of data produced by publicly funded research that are not shared and reused. This is despite the fact that Norway has endorsed the international FAIR principles – that data must be Findable, Accessible, Interoperable and Reusable. However, ensuring that data can be shared and used does not happen by itself. FAIR data requires both a physical data infrastructure and expertise. Assessing the quality of data requires specialist knowledge, and ensuring that data can be reused by both humans and machines may require expertise that not all researchers possess. The current solutions need to be developed into an ecosystem that meets all these needs.

This report proposes how such an ecosystem could be organised and funded. The report also identifies the benefits that can be realised if Norway sets ambitious goals in this area. The Committee proposes a framework that safeguards the need for a long-term perspective, coordination, quality assurance, and the establishment and development of data infrastructures for FAIR data. It also highlights which actors must be delegated responsibility for the different parts of the ecosystem.

Increased sharing and reuse of high-quality research data presents new opportunities, and our ability to take advantage of these opportunities will be a critical factor for participating at the forefront of research in the years to come. This means that the research system must be rigged in a different way than it was previously. It requires international standardisation, national coordination and prioritisation, as well as changes at the institutional level. International cooperation is more important than ever and is vital to ensure that the management of research data creates added value for research and society as a whole. Good infrastructures for the organisation of, access to and reuse of research data will be crucial in almost all areas of research going forward. This is, in turn, contingent on researchers having basic data management skills and the tools to manage data properly and securely at all stages of the research process. In addition to general competence-raising among researchers, more specialist expertise will be needed in data and data management both in and pertaining to research communities.

The report clarifies important features of development, as well as requirements for the research and innovation system that will enable us to achieve the level of ambition the Committee proposes:

- Researchers at Norwegian universities and research institutes must have access to data infrastructures that enable world-class research and education.
- By 2030, all subject areas in Norway must have access to expertise, guidance and curation of research data, either in the form of national solutions or by participating in European or international infrastructure cooperation, either in full or in part.
- In select areas, Norway must have world-class data infrastructures that become the preferred choice of international users.
- Norway must have a plan for how to escalate the organisation and funding of data infrastructures that make it possible to reap the benefits of the large amounts of publicly funded data that will be generated in the years ahead.

In order for the proposed level of ambition to be realised, the research budget must be sufficiently large to accommodate the necessary investments in data infrastructure and associated support functions.

The Committee also proposes how data infrastructures can be organised and funded. Among other things, it proposes that the relevant sector ministries must take greater responsibility and contribute more to the funding of data infrastructures. There is also a need for cooperation, and better coordination and division of tasks between the actors responsible for various data services. To conclude, the Committee presents a framework that supports more sustainable funding for establishing and operating data infrastructures. The framework is intended as a tool that the ministries, funding agencies and infrastructures can use to develop suitable funding models.

Background

In autumn 2022, the Ministry of Education and Research will present a revised Long-term Plan for Research and Higher Education. The aim of the revised plan is to give the research and innovation sectors a better tool to take on board the changes to the knowledge landscape, and develop policy in line with important national and international trends. The green and the digital transition are two such trends, and the research and education system of the future must be equipped to support data-driven research, administration, service development and value creation in a sustainable and ethically sound manner. Open research is a clear priority in Norway and internationally, and making research results available is a key element.

Remit from the Ministry of Education and Research

The Research Council was tasked by the Ministry of Education and Research with establishing a committee to draw up a background document containing recommendations on investments in infrastructures for FAIR research data and administration data that are particularly relevant to research. The recommendations and assessments presented in this report will be used as a basis for the new long-term plan drawn up by the Ministry of Education and Research.

The Committee's overall approach has been to assess possible ways of organising and funding an appropriate data infrastructure for FAIR research data in Norway. Important elements of this work have been:

- to identify which data services will be needed
- to propose who should be responsible for and manage different types of data
- to recommend what should be resolved locally, what should be organised as common national services, and what should be resolved through international cooperation
- to propose different investment and operating models
- to show the potential benefits and benefit realisation that a well-developed infrastructure for FAIR research data will bring about for generators and users of data
- to identify areas where new measures need to be implemented
- to recommend the level of ambition, including the ambitions Norway should have in the European cooperation in this area

The Committee held its first meeting in September 2021 and has held a total of nine meetings. Speakers from Norway and abroad were invited to several of the meetings to shed light on the subject from different perspectives. There have also been two input meetings, one with invited participants and one open for all, as well as dialogue with relevant actors during the work. In January, a memorandum was sent out requesting written contributions and the Committee received a total of fourteen responses. All the contributions have been assessed and taken into account in this final report.

The Committee members have experience of and expertise in various types of data infrastructures, and represent the higher education sector, the institute sector, the public administration and the business community:

Name	Position	Employer
Kenneth Ruud, chair of the committee	Chief Executive	Norwegian Defence Establishment

Ingrid Heggland, deputy chair	Senior research librarian, project manager	Norwegian University of Science and Technology (NTNU) Library
Ingunn Limstrand	Head of section	Norwegian Environment Agency
Klas Henning Pettersen	CEO	NORA, Norwegian Artificial Intelligence Research Consortium
Gunnar Bøe	Managing Director	Sigma2 AS
Tanja Storsul	Director	Institute for Social Research
Knut Kalgraff Skjåk	Head of Department	Sikt - Norwegian Agency for Shared Services in Education and Research
Christine Stansberg	Senior Advisor	University of Bergen, ELIXIR Norway
Pål Richard Romundstad	Professor	Norwegian University of Science and Technology (NTNU), Department of Public Health and Nursing
Alexander Refsum Jensenius	Professor	University of Oslo, Department of Musicology
Ann-Kristin Brændvang	Director of Department	Statistics Norway (SSB)
Carina Hundhammer	Head of Higher education and Research policy	Abelia
Geir Huse	Research director	Institute of Marine Research

The secretariat was made up of Siri Lader Bruhn and Rita Bergersen (head) from the Research Council and Katrine Weisteen Bjerde from the Norwegian Directorate for Higher Education and Skills.

Delimitation of the work

A committee of this size will not have experience and expertise from all areas of research and the public administration. It has therefore always been important to keep discussion to the level of general principles. This has resulted in overall recommendations rather than specific recommendations targeting particular areas.

Given the relatively short duration of the assignment, it has been necessary to make some delimitations to the extensive remit. The Committee has therefore not carried out its own surveys or analyses during its work. However, several other processes have recently been carried out that converge with the Committee's task. Assessments and recommendations provided during these processes have been part of the knowledge base and included in the Committee's discussions, where relevant. The following processes have been particularly relevant:

> ➢ *Requirements and funding strategy for national e-infrastructure for research in the period 2020–2030.* The working group's recommendations include increasing national investments in e-infrastructures to NOK 250 million annually, with a model for future adjustments of this investment requirement.
> ➢ It also recommends increasing investments to the Research Council – both to generic and to discipline and subject-specific national platforms – to at least NOK 400 million annually,

and that five per cent of national R&D should be invested in making research data available (letter from the Research Council to the Ministry of Education and Research 30 September 2021).

➢ *Datadeling i næringslivet* (2020) (Data sharing in trade and industry) A report from an expert group containing recommendations to the Norwegian parliament, the Storting, on how data from trade and industry can be shared: *rapport-fra-datadelingsutvalget2.pdf* ("Report from the Data Sharing Committee" – in Norwegian) (regjeringen.no).

➢ *Hvordan skal vi dele forskningsdata? Utredning og anbefaling om lisensiering og tilgjengeliggjøring*. (How should research data be shared? A study and recommendation on licensing and accessibility, The Research Council of Norway, 2021): *rapport_rettighets- -og-lisenssporsmal-ved-tilgjengeliggjoring-av-forskningsdata.pdf* ("Report and recommendations on licensing and accessibility of data" – in Norwegian) (forskningsradet.no).

➢ *Evaluation of the INFRASTRUCTURE initiative as a funding instrument* (Research Council of Norway, 2021). The evaluation committee recommends that the INFRASTRUCTURE initiative be continued as the main mechanism to support investments in national research infrastructures and Norwegian nodes in international research infrastructures: *infrastruktur-evalueringsrapport.pdf* ("infrastructure evaluation report" – in Norwegian) (forskningsradet.no).

➢ *Evaluation of UNINETT Sigma2* (2019) concludes that the infrastructure is important for Norwegian researchers and that activities, services and funding must be increased in order to meet the growing demands of an increasing number of users: Evaluation of UNINETT Sigma2 - Norwegian Agency for Public and Financial Management (DFØ) | kudos (dfo.no).

➢ *Felles infrastruktur og tjenester for FAIR forskningsdata (Common infrastructure and services for FAIR research data (FAIR report, 2022) – in Norwegian).* In its final report, the working group identifies major shortcomings in the current development of infrastructures, expertise and cooperation on FAIR research data: Open Science | *Felles infrastruktur og tjenester for FAIR forskningsdata* ("Shared infrastructure and services for FAIR research data" – in Norwegian).

In addition, the Committee mentions four processes that have commenced and are relevant to its work, but that have not yet submitted their proposals or reports:

• Legislative committee work: *Helhetlig regulering av videre bruk av offentlig informasjon (data)* ("Comprehensive regulation of wider use of public information (data)" – in Norwegian)

• Expert group, chaired by Simen Markussen: *Effektiv og sikker infrastruktur for deling og bruk av relevant statistikk og data i kriser* ("Efficient and secure infrastructure for sharing and using relevant statistics and data in crises" – in Norwegian)

• Expert group, chaired by Mari Rege: *Juridiske og etiske problemstillinger knyttet til innsamling, deling og bruk av randomiserte forsøk i kriser* ("Legal and ethical issues related to the collection, sharing and use of randomised trials in crises" – in Norwegian)

• Expert group, chaired by Heri Ramampiaro: *Deling av industridata* ("Sharing of industry data" – in Norwegian)

The structure of the report

Part 1 of the report describes the context in which the Committee's recommendations should be seen, and the infrastructure requirements that need to be met if Norway wants to provide for improved utilisation of Norwegian research data. In Part 2, the Committee reviews the benefits and costs associated with investing in data infrastructures, and recommends the level of ambition Norway should have. In Part 3, the Committee makes recommendations about the most cost-effective and sustainable means of organising and funding Norwegian data infrastructure. The report concludes with Part 4, which presents a framework for the funding of data infrastructure for research data and data of particular relevance to research.

PART 1

Infrastructure needs in a forward-looking and data-driven research and innovation system

The research system is presently undergoing two parallel transformations. Research is becoming more data-intensive and data-driven as a result of digitalisation and new technology, while at the same time, there is a significant shift towards open research, including the sharing and reuse of data. This results in less defined boundaries between different research disciplines, which in turn increases the opportunities for interdisciplinarity. It is in this way possible to achieve a more streamlined and quality-assured flow of data across the research sector, public administration and trade and industry. In this part of the report, the Committee assesses which changes are required for Norway to take a position at the forefront of research and contribute to precise and accessible knowledge in important areas of society. This forms the backdrop for the Committee's recommendations in parts 2, 3 and 4 of the report.

Terms used in the report

Research data:
Data that is generated or arises during the research process.

Public administration data of particular relevance to research:
Data from the public administration that is of interest for use in research.

Data infrastructure:
The tools, services and systems, including expertise, required to collect, analyse, store, organise, document and make available data.

Further in this report, for the sake of simplicity, we will:

- use the term *data infrastructure* as an abbreviation for *infrastructures for FAIR research data and public administration data of particular relevance to research*
- use the word *data* when we refer to both *research data* and *public administration data of particular relevance to research* as a whole, but clarify when it is specifically about research data or public administration data

1. FAIR data

Data sharing is nothing new, either within research or the public administration, but extensive digitalisation within most areas of society has provided more extensive opportunities for data sharing and in entirely new ways than previously. In order to take advantage of these opportunities, research processes need to change, and investments in data infrastructures and expertise that enable the sharing and reuse of data, must be prioritised.

Open data from the public administration has been a political guideline for many years. Data being *open* means, quite simply, that they are shared without any restrictions as to who can access them or for what purpose they can be used. In the Government's current strategy One digital public sector (2019–2025), the fundamental principle is that the public sector should share data when it can and protect data when it must. This is a fine and important principle, but a prerequisite for enabling the value of data to be utilised outside of the context in which they originate is that they are managed in a way that enables others, both humans and machines, to find, understand and use them.

In order to optimise the use and reuse of research data, the FAIR principles were presented in the article **The FAIR Guiding Principles for scientific data management and stewardship** in Nature in 2016. The acronym FAIR stands for Findable, Accessible, Interoperable and Reusable. These have become guiding principles in most fields of research around the world, and, in recent years, the public sector and private actors have also started incorporating them. The OECD, UNESCO and the EU have all endorsed research data being FAIR, and this is also a fundamental principle in the National strategy on access to and sharing of research data. All announcements in Horizon Europe require projects to have a data management plan that describes how the data management is in accordance with the FAIR principles, and the main goal of the partnership and the foundation European Open Science Cloud (EOSC) is to develop "a Web of FAIR Data and services". The Research Council has also introduced requirements for data management plans in research projects.

In other words, there is wide agreement and support for data being shared according to the FAIR principles. To achieve this in practice, however, such agreement must be followed up with necessary action. In the vast majority of cases, it is important to note that making data FAIR will be more

resource-intensive than making data open. Data infrastructures that facilitate such sharing will be absolutely essential, but so will increased expertise, improved coordination and division of responsibilities, both nationally and internationally.

1.1   FAIR data requires well-organised data infrastructures

FAIR data require data infrastructure that facilitates good data management. By data infrastructure is meant the tools, services and systems required to collect, analyse, store, organise, document and make available data. Realising FAIR data will require generic platforms for storing, archiving and making available large quantities of data, as well as discipline and domain-specific data infrastructures for different purposes.

Data infrastructures are needed that provide quick and easy access to data that comes with good descriptive metadata in accordance with international standards; methods for processing the data; systems for managing privacy; and intellectual property rights and licenses for further use, to make it easy to compile datasets from many different sources. The data infrastructures must work well individually, but different infrastructures provide functionality in different phases of the data's life cycle. Therefore, they must also be able to exchange information and work together so that users experience good and (almost) seamless functionality during all phases.

This will require data infrastructures that are designed for cooperation and knowledge-sharing across countries, disciplines and sectors. Data infrastructures designed for sharing and reuse of classified/confidential data demand additional requirements for data protection and secure data management. This applies, for example, to data that the nation has a responsibility to protect, and legal expertise and technology will be required to meet these particular requirements.

1.2   FAIR data generates a new demand for competence

As data becomes an increasingly central part of the research process, the research system must be retooled. This will create demand for skills at all levels, from basic FAIR data and data management competence among researchers to more specialist expertise in the support services and pertaining to the infrastructures that provide data services to research communities. The management of research institutions will also require a basic understanding of the necessary changes that must be made. The final report from the FAIR study (in Norwegian) highlights the need to raise skill levels in research communities and recommends measures are introduced to address this.

Operating a data infrastructure requires competent personnel with multiple forms of specialist expertise. Data stewards are employees with special expertise in data management. *A data steward assists researchers or others managing data to ensure good data management throughout all phases of the data lifecycle.*[1] This is a relatively new expertise group in the knowledge system, and increased demand for data stewards should be anticipated, corresponding to the increased demand for the reuse of data. It is difficult to predict the size of the demand for this type of expertise, but it has been suggested that up to five per cent of all research funding may have to be spent on data management going forward (Mons, 2020). The personnel of a domain-specific data infrastructure often need in-depth knowledge of the disciplines they are going to support, how data have been generated, which analyses are worthwhile performing on different types of data, and which types of metadata and standards are appropriate. At the same time, the personnel must have technical data skills.

---

[1]The description is taken from Open Science | List of terms for research data management (in Norwegian)

A number of data management tasks are so fundamental that all research institutions should ensure they have a certain level of competence. Other tasks, such as making a legal assessment as to whether a specific dataset can be shared openly, require specialist expertise that it may be prudent to hold in one or just a few communities. To provide researchers with a high standard of support and make the best possible use of resources, communication is essential in order to ensure that all research communities have access to the necessary expertise and that specialist expertise is not unnecessarily duplicated.

1.3   FAIR data depends on a clear division of responsibility

Investment in data infrastructures and competence-raising is expensive. It is therefore important to use resources in an efficient and sensible manner. An essential factor in achieving this is that cooperation is improved across infrastructures and institutions, and that roles and responsibilities are clearly defined. It will be important to establish which data infrastructure services should be offered at each institution, and which should be established as a national data infrastructure or a joint service. A clear division of roles and responsibilities will make it possible to avoid overlapping and unnecessary investments, and make better use of expertise across actors in the research system.

FAIR data will improve the flow of data within the research system, as well as the data flow between the research sector and other sectors. This is positive in itself but may present challenges in terms of where the responsibility for data infrastructures and management of data should lie. It will therefore be critical to clarify the division of responsibilities between the Ministry of Education and Research and other sector ministries that need research-based knowledge as a basis for policy-making and public administration.

1.4   FAIR data depend on international cooperation

International cooperation ensures research quality and renewal, and is vital to achieving FAIR data. International cooperation means that Norway can influence the development of standards for data management, metadata, methods and technology, and it also gives Norwegian research communities access to relevant data infrastructures in Norway and abroad. For many years, Norway has participated in international collaborations on data infrastructure and in international networks of expertise. In Europe, such collaboration in the future will largely take place through participation in the European Open Science Cloud (EOSC). Norway is now a member of the partnership, which brings new opportunities but also commitments. Through participation in EOSC, European researchers gain access to FAIR data and relevant services in virtual communities across scientific disciplines and national borders. The EOSC is developing something new, but is largely building on existing data and e-infrastructures (also Norwegian) for finding, accessing and reusing data. Participation in international collaboration means that the risks and costs are shared between the participating actors. This is in accordance with Norway's responsibility to contribute to international knowledge development and to help solve common problems and challenges, as described in Research cooperation – regjeringen.no.

1.5   FAIR data for better interaction between the public and private sectors

More and more research is being carried out in partnership between the public sector and business sector. This is an intentional political development, and making provision for cooperation will be important if Norway is to achieve its objectives of value creation, with data as a resource, in the new data economy (Data as a resource – Data-driven economy and innovation). Data infrastructures that

provide for more simple and flexible access to research data and public administration data give the private sector the opportunity to utilise these in business development and innovation. *Datafabrikken* (The data factory), which is under development, will help companies and public bodies with value creation based on data. This is an initiative that will build networks and expert communities in order to strengthen cooperation across sectors. The ambition is to reduce technical, legal and business-related barriers to sharing and reusing data. Gaia-X is a European initiative with representatives from trade and industry, politics and research from across Europe and the rest of the world, working together to build a distributed and secure data infrastructure. This infrastructure gives companies and citizens the possibility to exchange and share data and at the same time retain control over their own data.

2. What is needed going forward?

To succeed with FAIR data, better solutions must be available during the entire life cycle of data. To ensure the data is ready to become FAIR from the outset, more user-friendly support is required for planning data management with better support during the project phase. Not least, good solutions are needed to make it easier for both humans and machines to access archived data. To ensure that data will actually be used, good search services are also needed, enabling potential users to find available data from both research and the public administration. Many good projects are under way in this area at a European level, and it is important that Norway participates in and supports them.

Investments have been made in many different infrastructures in Norway in recent years, both through the Research Council's INFRASTRUCTURE initiative, through direct allocations from sector ministries and through local initiatives by research institutions. Internationally, investments have been made in research networks and data infrastructures, to which Norwegian research communities gain access through collaborations or membership. In addition, many Norwegian research institutions have gone far to facilitate FAIR data for their researchers by creating clear guidelines for data management, and by investing in stronger research support, local data services and the development of expertise.

It is worth noting that there is a lot of good data infrastructure in the public administration, which lays the groundwork for the use of public administration data in research. This data is collected by the public sector and stored in health, tax and labour market registries, the Brønnøysund Register Centre, NAV Norwegian Labour and Welfare Administration, and the National Archives of Norway, to name but a few. In addition, the ALM sector (ALM: archive, library and museum) has a range of data collections and databases that are purposed for reuse. In order to utilise data from the public administration in the future, it is crucial that the existing structures are able to further develop their services under good terms and conditions, while also facilitating the establishment of new infrastructures in areas where services are currently lacking. Taking full advantage of the potential value of data requires significant work on preparing data of the right quality and making them available to researchers and other actors.

There is still a long way to go before we have infrastructures in place for FAIR research data and public administration data of particular relevance to research in all subject areas. The FAIR study indicates that there is great variation in the organisation of infrastructures, and only a minority of data infrastructures are organised in a way that provides for interaction and data flow between infrastructures. A number of researchers feel that the provision of tools for managing research data is fragmented, and they would like to see greater cohesion between research infrastructures and data management services for the various parts of the research process.

The report from the expert group for [sharing of industry data](#) points to many of the same challenges as the FAIR study, and recommends improved coordination (especially at central government level) and competence-raising: "The state should ensure that government agencies contribute to cross-sector cooperation on data flow where this can create socio-economic value, regardless of whether this measure benefits the individual agency's core activities" (translation from Norwegian). They also point out that "a lack of expertise and a lack of focus on the need to process data in order to use advanced methods are significant barriers to data-driven value creation in the business sector and public sector". In relation to this, they also indicate that a lack of trust between actors is a barrier to sharing data, and that it is therefore important to strengthen expertise and capacity.

There is no comprehensive overview of which data infrastructures are in use in Norwegian research and public administration, and there are many different solutions for how they are organised and funded. In certain areas, we have come a long way in terms of infrastructure and expertise, and have a good division of responsibility with regard to organisation and funding, while much remains to be done in other areas. Going forward, it is important that Norway builds on what is working well, and finds sustainable solutions for organising and funding data infrastructures where they are currently lacking. The Committee goes into more detail about solutions in Part 3 of the report.

PART 2

Ambitions for data infrastructures in Norway

In this part of the report, the Committee proposes the level of ambition Norway should have for data infrastructures. It describes the benefits that can be achieved with a high level of ambition, and the risk Norway runs by choosing a level of ambition that is too low. Finally, it presents the outcomes for research and for society from choosing different levels of ambition.

> ## The Committee's proposal for the level of ambition for data infrastructures in Norway
>
> - Researchers at Norwegian universities and institutes must have access to data infrastructures that enable world-class research and education.
>
> - By 2030, all research fields in Norway must have access to expertise, guidance and curation of research data, either in the form of national solutions or, in full or in part, by way of participation in European or international infrastructure cooperation.
>
> - In select areas, Norway must have world-class data infrastructures that become the preferred choice of international users.
>
> - Norway must have a plan for how to escalate the organisation and funding of data infrastructures that make it possible to reap the benefits of the large amounts of publicly funded data that will be generated in the years to come.
>
> *In order for the proposed level of ambition to be realised, the research budget must be sufficiently large to accommodate the necessary investments in data infrastructure and associated support functions.*

3. Level of ambition for improved utilisation of data for research and societal development in Norway

The Committee believes that Norway must have a level of ambition corresponding to the overall goals set out in the Long-term Plan for Research and Higher Education 2019–2028. This means *enhancing Norway's competitiveness and innovation capacity, tackling major societal challenges and developing research communities of outstanding quality*. The possibility to collaborate and compete with the best communities internationally is crucial to developing outstanding research communities. The opportunities for new breakthroughs in research are limited, as are the opportunities to develop research on an international level, without access to data and the ability to link data. The level of ambition will have implications for the cost of future data infrastructure, but it will also affect what benefits it can bring to Norway.

4. Benefits

There is much to be gained from good management and sharing of data. Norway has a good basis for being at the forefront in this area, with the benefits it will bring to Norwegian research, public administration and business development. Data sharing requires effort and resources, but the benefits are increased value creation, more jobs and an efficient public sector that can offer citizens better and more coherent services. The European Commission has carried out studies that show that

the economic value of data from the public sector will increase from EUR 52 billion in 2018 to EUR 194 billion in 2030. This means that public-sector data has great value if it is properly managed and

> **Data for better public health**
> *HUNT has grown from what started as a single medical study to become a research centre and expert community that includes a biobank, data bank and infrastructure for sensitive data. The first Health Survey in the Nord-Trøndelag region, HUNT 1, was carried out in 1984–1986. Since then, more than 240,000 people from Trondheim have taken part in the surveys. HUNT research centre has a unique collection of health data and biological material.*
>
> *Long-term management in the research centre paves the way for important research. Every year, over a hundred research articles are published based on material from both internal and external researchers. The extensive dataset contributes to insight and new knowledge about public health and illnesses, genetics, lifestyles and environmental impacts.*

shared, and used for value-creating products and services.

4.1　Benefits for research

An obvious benefit of FAIR research data is that they can be made available for reuse by other researchers, who thereby avoid unnecessarily spending time collecting new data. This increases efficiency and leaves more time for data analysis and other research activities. In accordance with the principles of open research in general, FAIR research data will also have good metadata and method descriptions, providing for the reproduction and verifiability of the data. This makes it possible to correct errors, supplement existing data, improve methods and account for different studies producing different results. In other words, accessible research data will make research more efficient and help strengthen trust in the research. Qualitative research, which is not necessarily hypothesis-driven, will also benefit from being able to refer to structured and accessible background data. This could be relevant to the humanities and social sciences, where the sharing of, among other things, qualitative data can help provide support for arguments and conclusions.

> **Built-in privacy protection for efficiency and lower costs**
> *Through microdata.no, researchers gain affordable and immediate access to parts of Statistics Norway's source data, without the need to apply and without linking the data's use to defined projects. Researchers can use the service innovatively and exploratively to develop a project, and as a workspace and data warehouse when executing projects.*
>
> *Metadata is an integral part of the platform, and researchers can freely share datasets and analysis schemes. The ongoing project National Microdata Platform for Norwegian and International Research and Analysis is further developing microdata.no into a generic distribution platform for data from any data owner. The Cancer Registry of Norway is a pilot partner in the project. This is possible because microdata.no has embedded privacy protection. The users work on a metadata level, without seeing individual data, and the data never leaves Statistics Norway. All results go through a set of confidentiality filters making them anonymous. Any attempts to circumvent confidentiality are blocked by the built-in privacy protection.*
>
> *The Fafo Research Foundation report Seniorer i skolen (2022:05) ("Seniors in school" – in Norwegian) examines age, gender, working hours, type of school and retirement age for teachers during the period from 2010 to 2019.*
> *The data is taken from microdata.no. Statistics Norway has calculated that a similar conventional loan of data would cost around NOK 30,000–40,000 and have a production time of two to four months, depending on the current demand. This does not include the researchers' costs and the time spent preparing an application.*

(Skrivefeil I kildeteksten: 'helt' istedenfor 'helst')

The reuse of data from the public administration in research will present enormous opportunities, in areas such as population, health, transport, natural resources, the climate, language etc. Norwegian register data, for example, are in a class of their own, and there is a great deal of untapped potential in using them in research. Better provision for the reuse of public administration data in research will provide knowledge about complex societal conditions that cut across research fields and topics.

Results from the research will in turn provide valuable insight and knowledge that can provide a better and more knowledge-based basis for development and decision making in the public administration.

4.2   Benefits for the public administration, society and citizens

The ability to collect and process data from various sources provides a better decision-making basis and more reliable predictions within important areas of public administration. Data that the public sector itself collects, or that are used as a basis for analysis in research, provide an important knowledge base for planning and decision-making that affect societal development in important areas. FAIR data also allow sector boundaries to become less defined, making it possible to offer more comprehensive and user-friendly services to citizens, cf. the overall strategy for cross-ministerial coherence Gode hver for oss. Best sammen 2021–2025 (“Good alone. Best together 2021–2025” – in Norwegian).

It is important that research results are used in society, as a basis for media cases, service development, policy development and business activities. Transparency is essential when it comes to the data basis for research results to instil people's trust in the research and to be able to discuss it across interests and values. Indirectly, this contributes to better research and thus a good basis for decision-making. A well-documented knowledge base, stemming from the systematic use of research data in areas such as health, the economy, labour market, weather, climate, traffic and air quality, allows people to make informed choices based on their own risk assessment. This contributes to an open and well-informed democracy.

Useful data services for the general public

- *yr.no – a data service from the Norwegian Meteorological Institute and NRK, the Norwegian Broadcasting Corporation*
- *The Norwegian Biodiversity Information Centre - a bank of knowledge about biodiversity*
- *Dictionaries and other language resources through CLARINO*
- *Libraries, museums and archives*
- *helsenorge.no/en*
- *Statistics, research and analysis from Statistics Norway*
- *Encyclopaedias, such as Great Norwegian Encyclopedia (in Norwegian)*

4.3   Benefits for trade and industry

Public-sector data provides opportunities for innovation and creativity in trade and industry, including the development of new services, products and business models. Collaboration on infrastructure and the use of data streamlines value chains in the business world. Its benefits include reducing faults and errors and eliminating work tasks and intermediaries. Large and preferably open databases are

---

From the soil to the clouds
DigiFarm is a Norwegian start-up that helps farmers and agricultural enterprises to increase their yields and reduce the costs of fertilizers and crop protection. Using deep neural networks (AI) based on high-resolution satellite data, DigiFarm has developed a fully automatic and standardised method for detecting and updating soil boundaries and sown areas for each growing season. This can replace manual, imprecise and time-consuming tools that farmers currently use for grant applications, soil samples, fertilizer plans, etc. At the authority level, it automates what previously entailed manual verification of production subsidy applications.

a prerequisite for using artificial intelligence, which allows the possibility of advanced analyses. Such open databases also provide great opportunities for value creation, not least for start-ups. In an increasingly data-driven economy, smart compilation and analysis of large amounts of data may improve economic competitiveness and innovation. This is supported in *Report No 22 to the Storting (2020–2021) Data as a resource. The data-driven economy and innovation* , which highlights the benefits of sharing data, including to strengthen the competitiveness of business both nationally and internationally.

4.4   International cooperation is crucial to reaping the benefits

Through a number of strategic investment initiatives and new regulations, the European Commission has, since 2014, worked on the goal of facilitating a more balanced and targeted flow and use of data across sectors. Norway's participation in this research and data collaboration provides benefits in the form of easier access to data from other contributors, improved analytical results and new digital services. Initiatives that have promoted the development of a robust European data economy include the regulation on the Free Flow of Non-Personal Data, the Cybersecurity Act and the Open Data Directive. Recently, the Commission presented a new proposal for a regulation on harmonised rules on fair access to and use of data (the Data Act). The regulation contains a framework for cross-sector management of data that will ensure users retain access to their own data. This promotes non-personal data sharing and security measures to prevent third countries from gaining access to EU data in the absence of an international agreement.

Although Norway is not part of the EU, we have access to both the EU's data market and European funding initiatives such as Horizon Europe and the Digital Europe Program (DIGITAL). It is here that Norway can coordinate national investments in data infrastructures with European investments for the development of cross-sectoral data infrastructure. Participation in international data infrastructures gives Norwegian actors the opportunity to realise the benefits of improved flow and use of data across sectors in Norway and other European countries. Participation in international data infrastructures is a prerequisite for innovative research of a high international standard. The EU has also invested large sums of money in data analysis and computing capacity through EuroHPC.

Artificial intelligence contributes to world-leading research

Understanding protein function is essential to all life. A protein's shape determines its function. In the past, experiments to find a protein's structure were very difficult to perform and were therefore expensive. It is not unusual for advanced methods to take a year to carry out, and it can normally cost up to NOK 1 million to determine the structure experimentally.

In 2020, there was a very special development in this field. The company DeepMind developed the AI algorithm AlphaFold, which, based on data from the Protein Data Bank, quickly predicts the shape of completely new proteins. The algorithm quickly mapped the entire human proteome (all proteins in the human body) and now other species are successively being mapped.

In 2021, AlphaFold was named "Method of the Year" by the journal Nature. DeepMind opened the algorithm, and on the initiative of NORA – Norwegian Artificial Intelligence Research Consortium, USIT at the University of Oslo (UiO) and Uninett Sigma2, the algorithms were quickly incorporated in Norwegian infrastructures and made accessible to Norwegian research communities. The programme is available to the entire research community and can, among other things, be used to develop medicines.

---

5. Costs of investing in infrastructure for FAIR research data and public administration data

Below, the Committee describes the costs that must be anticipated in a data infrastructure that provides for FAIR data. An assessment of the financial implications and benefits of having a high level of ambition is included at the end of the chapter. Within the framework of its work, it has not been possible for the Committee to produce reliable figures for the cost of Norway's current infrastructures.

5.1   Costs of e-infrastructure that is a prerequisite for data-based research and public administration

E-infrastructures are ICT-based infrastructures that are crucial for data-driven research, such as high-performance computing and storage of big data during the research process. Such infrastructures are not directly related to FAIR sharing of data, but are just as much a prerequisite for research, analysis and data mining – extracting data from existing databases to generate new information. This type of infrastructure competes for the same funding as other types of data infrastructure. That is unfortunate because they tie up large amounts of competition-based funding. The Committee therefore supports the Research Council's recommendation to the Ministry of Education and Research that such fundamental data infrastructure should be funded directly and not competition-based. The recommendation from the evaluation of the INFRASTRUCTURE initiative also concludes that these are fundamental infrastructures that should not be competing for funding. The Committee will return to this point in Part 4.

Research infrastructures that require regular large investments and that have system-wide effects should be funded centrally and not through the INFRASTRUKTUR competition-based funding scheme. Central funding will ensure continuity of its system-wide importance and to avoid difficult priorities having to be made between these infrastructures and more specialised infrastructures (2021:8).

High equipment costs are associated with e-infrastructures that provide data services for high-performance computing, and storage of large amounts of data. Among other things, the investment requirement is shown in the 2019 report *Behov og nasjonal finansieringsstrategi for nasjonal e-infrastruktur for forskning for perioden 2020–2030* ("*Needs and national funding strategy for national e-infrastructure for research for the period 2020 – 2030" – in Norwegian)*, and the Committee supports these recommendations. The need for investment is increasing in parallel with the demand for services related to data storage and high-performance computing. This development can be seen in applications to the Research Council and the estimated budget for upgrading equipment for Sigma2, which is around NOK 200 million per year. This illustrates the need to increase the budget considerably. It is important both to further develop technology and to meet the needs of various expert communities and sectors.

5.2   Data infrastructure costs for FAIR data

The cost elements in a data infrastructure are services that manage data throughout the research life cycle – before, during and after a research project – and ensure that the research is carried out technically, methodically, legally and ethically, and maintains high quality. It is also necessary to finance the work that must be done beyond individual research projects, such as the development and maintenance of standards, and the management and long-term preservation of data after a project has ended.

As previously stated, there will be an increased need for skilled personnel to safeguard an up-to-date data infrastructure in the longer term. Among other things, increased costs should be expected in relation to competence-raising among students and researchers, through first degrees, continuing and further education, and research training programmes. The Committee believes that the responsibility for these costs must mainly be covered by the research institutions themselves, and that they must be integrated into regular research education. It is equally important to take this into account when calculating costs. This is also emphasised in the FAIR study:

> *The study revealed that a lack of expertise and a culture for sharing and reusing research data is probably the biggest obstacle to advancement. Covering the need for competence-raising and guidance for all researchers in Norway demands both significant investment of resources and good coordination of competence-raising measures among local, institutional and national research communities. In general, competence must be raised at all levels in the research system, and available capacity for guidance and operational data management must be increased significantly (2022:4).*

The FAIR study documents great variation when it comes to how far Norwegian research institutions have come with respect to sharing research data. The Swedish Research Council makes the same point in the status report on FAIR data (in Swedish with summary in English) and makes the following recommendations about the need for investment:

- *A specific financial investment is made to increase the infrastructural support for the national work on open access to research data at higher education institutions, public agencies and infrastructures.*
- *The Swedish Research Council is awarded specific funding to support activities that promote open science nationally and internationally, for example through participation in the European Open Science Cloud (EOSC).*

It is important to take the costs of international participation into account when developing data infrastructures. Norway's partnership in EOSC includes a commitment to contribute EUR 500 million in in-kind contributions during the period 2021–2030. This can be in the form of contributions to

activities funded by the Commission or to [activities](#) that are in line with the [strategic agenda](#) of the EOSC.

The need for FAIR research data services is growing. It is difficult to quantify what this need equates to monetarily, but an interesting source of information is the growing number of applications for such infrastructures, including to the Research Council's INFRASTRUCTURE initiative. Since the initiative was established in 2009, the proportion of applications for funding for data infrastructures, which fully or partially relate to tools for data management, has increased. In the last call for proposals in 2020, around 40 per cent of the projects applied for funding for data management (NOK 3.3–3.7 billion of the total amount applied for, NOK 9.3 billion). This demonstrates that a large proportion of the indicated need for data infrastructure funding is not covered by competition-based funding.

Institutions themselves have made significant efforts in the development and operation of data infrastructures. A review of the data infrastructure projects granted through the INFRASTRUCTURE initiative shows that the initiative has provided support for 66 per cent of the total project costs, and that the institutions themselves have largely funded the rest (see table).

Table 1 Overview of investments through the Research Council's INFRASTRUCTURE initiative during the period 2009–2018.

Data infrastructure projects per area	Total sum (million NOK)	Contract sum with the Research Council (million NOK)
Natural sciences	942.6	609.3
Health and personal data	532.1	488.8
Humanities	316.9	214.4
Social sciences	145.5	107.3
Generic	1,321.0	784.3*
TOTAL	**3,258.1**	**2,204.1**

*Includes NOK 237 million allocated under item 53 to Sigma2 in 2019

The findings from the FAIR study, the evaluation of the INFRASTRUCTURE initiative and evaluations from the Swedish Research Council show that it will be resource-intensive to develop and operate data infrastructures with the services that researchers and other users need. In contrast to investment in equipment, where the bulk of the costs come at given times in relation to procurement and upgrades, establishing a data infrastructure will tie up significant personnel resources throughout the infrastructure's life cycle. In other words, it is important to have a long-term perspective when calculating the costs of a data infrastructure.

As mentioned previously, the European Commission estimates that it is necessary to spend five per cent of total national research funding on data management. For Norway, this will equate to about NOK 2 billion, if based on the [state budget analysis for 2021 by the Nordic Institute for Studies in Innovation, Research and Education (in Norwegian)](#). Proposition No 1 to the Storting (Resolution) (2020–2021) estimates total allocations for R&D activities to be NOK 40.9 billion. The recommendation that five per cent of national R&D should be invested in research data is based on a [calculation carried out by PwC EU Services](#) in 2018 on behalf of the Directorate-General for Research and Innovation. PwC estimated the actual loss resulting from the FAIR principles not being implemented in Europe at between 3.4 per cent and 8.7 per cent of the total R&D investment.

6. Step-by-step realisation of FAIR data

As shown above, it is expensive to make data FAIR, because it relies on both specialist expertise and investment in data infrastructure. When seen in conjunction with the large amounts of data collected, it is necessary to assess and prioritise which data should or can be made FAIR. Hopefully, good, large capacity infrastructures will eventually be in place in all research areas, and what is currently considered specialist expertise will become elementary knowledge, at all junctures. The need to assess and prioritise could then possibly be reduced and all data could in principle be made FAIR. Before arriving at that point, it makes sense to consider a step-by-step realisation of FAIR data.

Some data are easier than others to make FAIR because they require less adaptation to be shared and reused, while others require a lot of processing and/or protection. A number of disciplines have also come a long way towards making data FAIR, by standardising metadata, formats, ontologies and processing, while others have yet to cross the starting line. Instead of aiming for all data to become FAIR at once, it perhaps makes sense to first harvest the "low-hanging fruit" before focusing on the more complicated datasets and disciplines that are not yet ripe. This could make it possible to achieve greater transfer value, both in terms of technical solutions and expertise.

Improved public health effects from the improved use of data

The societal costs associated with disease in the form of lost quality-adjusted life years (QALY) were calculated by the Norwegian Directorate of Health to be over NOK 1,400 billion in 2013. A one per cent reduction in lost QALY resulting from the improved use of health data gives a socio-economic effect of NOK 14 billion. Improved use of health data could lead to more efficient operation of the health service, which makes up a significant part of public expenditure. However, realising these benefits will require a completely different pace of innovation than we have seen until now. It requires a strong health ecosystem to be established: To succeed would require working closely with academia, the health industry and the ICT industry (Menon Economics, 2018:3).

Furthermore, some data potentially have great value for further utilisation, while others will have limited value beyond the research on which they are based. Any data with potentially great value should then be prioritised before data assumed to have limited value. It may be difficult to assess and predict the future value of data because they may be utilised in completely different disciplines, industries or areas of society than those in which they were obtained. The main principle should nevertheless be that research communities make the assessment themselves, and preferably in dialogue with other relevant disciplines and other actors that may be potential users.

Another way to realise FAIR incrementally is to make different requirements of each letter of the FAIR principles. It is often much more resource-intensive to make data interactive/interoperable and reusable (I and R) than to make it findable and accessible (F and A). Just making the data findable could be hugely beneficial. In the first instance, provision could be made for all datasets to be findable, through machine-readable data management plans, and registration in data catalogues and registers, so that potential users know where the data is and how they can access them. So initially, the "low-hanging fruit", and datasets with high potential value for reuse, can be prioritised using the principles I and R.

Regardless of which steps are chosen to realise FAIR data, it requires a coordinated national and international effort, technical facilitation and standardisation, and the development of expertise. The Committee goes into more detail about recommendations on how this can be resolved in Part 3 of this report.

7. Assessing ambitions and risks

The Committee is of the opinion that the level of national investment in infrastructure for data must be in accordance with the expectations of the European Commission and Horizon Europe with respect to open research requirements. An important consideration for the investment is that it will give researchers at Norwegian universities and research institutes access to infrastructures that make world-class research and education possible. At the same time, the Committee believes that users in the public and private sector should have access to data infrastructures that provide easy access to high-quality research data when needed in planning and decision-making processes relating to societal and business development. It is essential that the costs of data management are covered because, in many cases, the benefits will be realised in other parts of the ecosystem than where the collection and provision for reuse of the data took place. The Committee will return to this point in Part 3.

If technological opportunities are not utilised, Norway's competitive advantage – in digitalisation and data registers with the potential to link data from different sources – is in danger of waning. Long-term management of data requires both technical expertise and capacity, and adequate resources for upgrading and modernising technical data infrastructures and data inventory. If such costs are not covered, upgrades and development will be impeded, and there is a huge risk that solutions become outdated, infrastructures degrade and long-term investments are lost. The updating and development of systems is a prerequisite for Norwegian and international research communities to utilise data collections for the benefit of society. The risk of not investing in forward-looking data infrastructure is that Norwegian research communities will not participate in the development of knowledge from international cooperation, and that Norway will lag behind as a knowledge nation. In all parts of society, it is important to have a reliable data basis when making assessments and decisions that have consequences for societal development. We have recently seen examples of how important it is to quickly access data for decisions that affect people's lives and health.

- The consequences of setting the level of ambition too low will be that Norwegian research communities risk losing the opportunity to participate at the forefront of international research. This will result from lack of investment in good and forward-looking infrastructure for FAIR research data and public administration data of particular relevance to research.
- There is a risk that the benefits of the data infrastructure services will not be fully reaped if personnel and competence-raising costs are not taken into account.

The EU publication *Cost-benefit analysis for FAIR research data: Cost of not having FAIR research data 2019* also indicates which values are at risk by not implementing FAIR data:

> *FAIR data has a direct and quantifiable benefit on research and the knowledge economy. A recent study by the European Commission and PwC noted the cost of not making Europe's research data FAIR to over €10 billion per year.*

8. New funding or investment in the existing framework?

As shown above, investment in establishing and operating data infrastructures carries large costs, but will yield large benefits in the long term. The investments should therefore be assessed from a cost-benefit perspective. The advantages will often be considerable, and both short-term and long-term benefits from implementing FAIR data must be substantiated when investments are being made. The illustration below (table of scenarios on page 27) shows the benefits and risks of adding new funding and increasing investments (scenario 1) versus reprioritising within the current

framework (scenario 2) versus neither adding new funding nor reprioritising (scenario 3). The costs that must be covered to implement FAIR data require the reprioritisation of funding within the current budget or a budget that is higher than the current level.

The risk of not sufficiently investing is that the costs accumulate while the benefits do not materialise. If Norway is to continue to cooperate in international research, we must be qualified to do so and prioritise funding accordingly. It is essential to invest more in Norway's data infrastructures. By not investing in a good framework for data sharing and use, we miss out on knowledge and technological development in areas that are important for developing solutions for trade and industry and society as a whole. Norway should also aim to drive development in certain areas.

The Committee's assessment

The Committee is of the opinion that only scenario one will fulfil the level of ambition proposed at the beginning of this section. Scenario 2 risks other parts of the research sector being deprioritised, or possibly organised in ways that could reduce the quality of research. The Committee does not consider scenario 3 to be a viable alternative if Norway is to have high ambitions for Norwegian research.

Table of scenarios

Scenarioer	Investeringsmuligheter	Forskerårsverk	Forskningsaktivitet	Kvalitet i forskningen	Verdiskaping utenfor forskningssektor
Scenario 1: **Friske midler til datainfrastruktur**	Norge får mulighet til å investere i nødvendig framtidsrettet datainfrastruktur.	Investering i datainfrastruktur påvirker ikke andre deler av forskningsbudsjettet, og antall forskere kan forbli som i dag.	Framtidsrettet datainfrastruktur vil gi forskere tilgang til en helt annen mengde data og på en helt annen måte enn i dag og gi mulighet for å øke forskningsaktiviteten.	Norske forskere får tilgang til datainfrastruktur som gjør at de kan ta del i den datadrevne forskningen og dermed være en del av forskningsfronten.	Gode datainfrastrukturer vil gi andre sektorer og næringsliv store muligheter til å skape verdier basert på forskningsdata.
Scenario 2: **Omprioritering – mer av forskningsbudsjettet settes av til datainfrastruktur**	Norge får mulighet til å investere i nødvendig framtidsrettet datainfrastruktur.	Mer penger til investering i datainfrastruktur vil gå på bekostning av andre deler av forskningsbudsjettet, og antall forskere må kanskje reduseres.	Framtidsrettet datainfrastruktur gir forskerne tilgang til data på en helt annen måte. Dette vil øke forskningsaktiviteten på viktige områder. Færre forskerårsverk og mindre ressurser vil imidlertid gi mindre forskningsaktivitet på andre områder.	Norske forskere får tilgang til datainfrastruktur som gjør at de kan ta del i den datadrevne forskningen og dermed være en del av forskningsfronten – men reduksjon i andre deler av forskningsbudsjettet kan likevel føre til lavere kvalitet.	Gode datainfrastrukturer vil gi andre sektorer og næringsliv mulighet til å skape verdier basert på forskningsdata – men lavere kvalitet kan føre til mindre verdiskaping.
Scenario 3: **Ingen omprioritering – potten til data-infrastruktur økes ikke**	Norge får ikke mulighet til å investere i nødvendig framtidsrettet datainfrastruktur.	Investeringer i infrastrukturer for datainfrastruktur påvirker ikke andre deler av forskningsbudsjettet, og antall forskere kan forbli som i dag.	Norske forskere vil ikke få tilgang til framtidsrettet datainfrastruktur og kan derfor ikke ta i bruk data på nye og mer effektive måter. Dette vil etter hvert føre til mindre forskningsaktivitet sammenlignet med andre land som har tilgjengelig datainfrastruktur.	Norske forskere vil raskt falle bakpå uten tilgang til datainfrastruktur som internasjonale kollegaer har, og vil ikke være i stand til å være en del av forskningsfronten.	Mangel på gode datainfrastrukturer vil gjøre det vanskelig for andre å gjenbruke disse dataene til verdiskaping utenfor sektoren.

Grønne felter: Situasjonen forbedres fra i dag ¤ Gule felter: Det er risiko for forverring av situasjonen ¤ Røde felter: Situasjonen forverres fra i dag

PART 3

Sustainable organisation and funding of data infrastructure

Long-term and sustainable solutions are required to ensure Norway gets the most out of its investments in data infrastructure. This means clarifying responsibility, drawing up priorities, coordinating overlapping data infrastructures and setting up good models for funding establishment and operation. This part of the report provides input to the Committee on how to address this, and it can be seen as a basis for the framework for funding of data infrastructures proposed in Part 4.

The Committee's recommendations on funding and organisation of data infrastructures

Funding and cash flows

1   The Ministry of Education and Research must assume overall coordination responsibility for funding national data infrastructure and set aside sufficient funding for this budget item.

2   The sector ministries' research budgets must include both direct and indirect costs for data infrastructure in the areas in which they need research and knowledge.

Coordination and division of responsibility

3   The research communities must assume a special responsibility for identifying and communicating data infrastructure needs. This includes highlighting potential benefits.

4   The research institutions are responsible for facilitating FAIR sharing of data wherever required. This can be done through adapted institutional services or access to national or international data infrastructures. Government institutions have a particular responsibility in this area.

5   Schemes should be established for encouraging collaboration on infrastructure and transfer of expertise between relevant communities in the form of networks, meeting places and/or centres.

6   Any decision to establish and further develop national data infrastructures should be made at central government level either via the Research Council's INFRASTRUCTURE initiative or through other mechanisms that ensure an assessment of the infrastructures' place in the national and international data infrastructure landscape.

7   Efforts should be made to facilitate participation in international data infrastructure collaboration to ensure efficient use of resources and further development of Norwegian nodes and pertaining expert groups.

8   The sector ministries must take responsibility for making relevant public administration data available for research.

9   Arenas must be established for prioritising which data collections and data infrastructures resources should be invested in making FAIR. This must be done together with the research communities, public administration, trade and industry and other relevant actors in society.

9. Sources of funding and cash flows

The Ministry of Education and Research is responsible for a well-functioning research system, which includes being responsible for research infrastructures. As shown in Part 1, more data-driven and data-intensive research will bring about an increase in the need for data infrastructures. Part 2 shows that investments in FAIR research data greatly benefit all segments of society. The Committee considers it vital to take a whole-system approach to sources of funding and cash flows. The benefits that we reap are a common good for Norway, and we must therefore shoulder the investment needed together. Achieving the ambitions for data infrastructures will require a clear division of responsibility also in terms of funding. The *Veileder for sektoransvaret for forskning* ("Guide on sector responsibility for research" – in Norwegian) (2017) underlines that each ministry is responsible for research within its areas of responsibility, and that they in this way help to implement Norway's common research policy. The guide does not specifically state that this concerns research infrastructure, but the Committee believes this is a given in that this type of infrastructure is by and large a precondition for carrying out good research.

9.1   The role of the Ministry of Education and Research

The overall responsibility of the Ministry of Education and Research to ensure a well-functioning research system also entails a special responsibility for coordinating and funding research infrastructure. Funding under item 54 in the Ministry's budget is to be spent on "relevant and updated research infrastructure of national, strategic interest that supports high-quality research for an innovative and sustainable society". This funding finances the Research Council's INFRASTRUCTURE initiative, as well as Norwegian contributions to selected international infrastructures and the basic allocation to Sigma 2 AS. The allocations under this item have increased steadily over the last few years to address the major research infrastructure needs, and an increasingly large proportion of the funding under this item goes to data infrastructures. It is vital that this budget item is not reduced if we are to achieve the level of ambition for reuse and further use of data in Norway.

> **The Committee recommends:**
>
> *The Ministry of Education and Research must assume overall coordination responsibility for funding national data infrastructure and set aside sufficient funding for this budget item.*

The block grants to universities and university colleges, which are also funded over the Ministry of Education and Research budget, gives the institutions the strategic scope to use the grants in line with sector goals and their own goals. The Committee believes it is important that the Ministry's management dialogue with the institutions highlights their responsibility for helping to achieve the goal of FAIR sharing of data, and thus ensure that researchers get the support they need to make data FAIR.

9.2   The role of other ministries

Through their sector responsibility for research, the ministries are responsible for, among other things, contributing to research and competence-building in the sector, ensuring that research is carried out for policy-making and public administration, and facilitating high scientific quality and relevance in research. It is important to consider this responsibility in context with the need to invest in cross-sectoral research infrastructure.

Many research infrastructures are currently funded by the state through ministries and directorates. This is done when the sector ministry needs to establish and operate a data infrastructure in a

specific area. The funding generally takes the form of a management assignment to a research institute or as a grant to a public administration institute. In the event of a management assignment, the ministry or directorate finances the development and operation of data infrastructure that provides a basis for the research institute's administration of data for the state. Infrastructure run by NILU is one example of funding in the form of a management assignment. However, when grants are awarded to public administration institutes, the allocation for development and operation of infrastructure often accompanies the allocation letter from the ministry in question. One example here is the Institute of Marine Research, which receives such allocations from its owner ministry.

Some sector ministries award large sums to research and, by doing so, generate a lot of data that often needs computing power and data infrastructure for storage, access and archiving. In some cases, the allocation includes funding for data infrastructure, as described above. In other cases, funding is allocated to the research activity alone, for example allocations through the Research Council and other funding agencies. In the latter case, it is often problematic that the sector ministries see the research activity itself as their responsibility, but do not take into account the consequences of this for the national data infrastructure. This is challenging because the research ties up capacity in the infrastructures without them receiving extra funding. Based on each ministry's sector responsibility for research, it is reasonable to expect that they help to finance data infrastructure beyond the direct investments they make in connection with the management assignments and grants to institutes. In the areas in which the sector ministries finance data-intensive research, they must also provide funding for the data infrastructure required to enable them to meet their knowledge needs, which often cut across disciplines and sectors.

There are also examples of infrastructures that are found in the borderland between research, public administration and dissemination, such as in the archive, library and museum sector. Dissemination is often the primary activity of such infrastructures, which can often make it challenging to facilitate good use of research. Considerable benefits can be envisaged here through better interaction with the libraries and museums run by the universities.

> **The Committee recommends:**
>
> *The sector ministries' research budgets must include both direct and indirect costs for data infrastructure in the areas in which they need research and knowledge.*

10. Division of responsibility and coordination in a complex landscape of data infrastructures

As described in the FAIR study and in Part 1 of this report, the current status is that there are a range of different data infrastructures (and pertaining services) of different sizes and with different areas of impact. Some infrastructures are local, some are national, while others are international. Some data infrastructures are generic and can be used in a range of disciplines, while others are adapted for use in specific disciplines. Others, in turn, can be purposed for broad dissemination tasks (such as libraries and museums). Data infrastructures are generally established to meet an existing need, and such needs have often arisen in several places and within several disciplines at the same time. This results in partially overlapping data infrastructures in some areas. In the meantime, there is a lack of data infrastructures in other disciplines, because they have not traditionally managed such large quantities of data, or because there are no incentives or culture for sharing data.

In the wake of the authorities and funders requiring that data be made FAIR, a need for data infrastructures has also emerged in these areas. There is no complete overview of the landscape of

data infrastructures used by Norwegian research communities, nor have the needs going forward been mapped in any detail. A survey of the research infrastructures and services in common use among Norwegian researchers today was initiated as part of the FAIR study. It provides a good overview of generic data infrastructures, but does not go into the individual disciplines in any detail. The results of this survey have been published on the website openscience.no.[2] Sikt is continuously developing the website in collaboration with the Research Council, Universities Norway and the institutions.

The FAIR study provides an indication of what works and which areas will be particularly important to follow up. It also sets out that considerable funding will be needed to achieve better functionality for services that require capacity and expertise (2022:39). This applies in particular to data that cannot be openly shared and must be protected.

Achieving cost-effective investments that cover the data infrastructure needed is contingent on a clear division of responsibility and coordination. Given the huge diversity of data, with different properties, complexity and management needs, the Committee does not consider it expedient to propose a "super infrastructure" to meet all needs. It is nevertheless important to promote an organisation that will help to achieve more coordinated data infrastructures than we have today, and that they together meet the overall need and support interdisciplinary research and better flow of data. We need a better division of labour and coordination than is the case today, and an overview of the responsibilities and roles that different actors have for making data FAIR.

Responsibility in this context often implies a responsibility for funding. This does not mean, however, that the responsible party alone should carry the costs of establishing and operating infrastructure. We will return to various funding models in Part 4, where we propose a framework for funding of data infrastructure.

10.1   The research communities' responsibility

The research communities know what their data infrastructure needs are, both for being able to carry out high-quality research and for meeting requirements and guidelines on making research data FAIR. To enable institutions, funding agencies and the authorities to make the right decisions on investments in data infrastructure, it is important that needs are identified and clearly communicated by the research communities. It will be useful when prioritising between investments that the potential benefits are evident, and that communities with concurrent needs join forces to provide input to decision-makers. This could take the form of coordinated applications for competition-based funding.

> **The Committee recommends:**
>
> *The research communities must assume a special responsibility for identifying and communicating data infrastructure needs. This includes highlighting potential benefits.*

10.2   The institutions' responsibility

During the life cycle of research data, researchers need different data infrastructures and pertaining services. Some of these data services must be perceived as "basic" data infrastructures, because they are services that all institutions must be expected to offer their employees. To the extent institutions do not offer such services themselves, they must be expected to enter into agreements with other

---

[2] See articles at openscience.no about the different types of infrastructures and tools, Open Science | Åpen forskning

actors that can do so. Many research institutions develop their own, more specialised data infrastructures, and several of them are also offered to researchers at other institutions. It is important in such case that the institutions are aware of any overlaps with other institutions' services and what national data infrastructure is available, and do not duplicate these unnecessarily. The management of research institutions must also make it their responsibility to ensure that staff have access to national data infrastructure. This entails among other things setting aside money for any user charges in research projects that the institutions fund over their ordinary budgets. It is also important to think in a life cycle perspective when making decisions. Many researchers are coaxed into using simple, free commercial services that are available with very unclear time horizons and a lack of national alignment.

> **The Committee recommends:**
>
> *The research institutions are responsible for facilitating FAIR sharing of data wherever required. This can be done through adapted institutional services or access to national or international data infrastructures. Government institutions have a particular responsibility in this area.*

10.3 Encouraging networking and knowledge-sharing

Local knowledge is an important factor for enabling researchers to use data infrastructure and pertaining services in an expedient manner. Research administration, research libraries, IT departments and local research support play an important role as the first line for researchers who are unsure of how they can and should manage data. However, it is neither cost-effective nor expedient for each individual research institution to have their own specialist expertise in all areas of data management and data administration. The FAIR study sets out that better access to knowledge and guidance is needed in specialised areas that many of the institutions do not have the capacity to provide. It may therefore be sensible to share the responsibility for different areas of expertise.

Obtaining external services is particularly relevant for small institutions that do not have well-developed support functions, e.g. their own library services. They can benefit greatly from collaborating with large research institutions and national actors by helping to strengthen existing national and international networks and, if relevant, establishing new ones. It may also be relevant to consider establishing centres of expertise and/or expertise hubs for developing and using this knowledge to make data FAIR. Such expert groups or centres in important domains can provide guidance and advice and help develop knowledge where legal, ethical, organisational, semantic and technical expertise is needed. They must work closely with the infrastructures and research communities. Overall, this will make good use of the expertise, promote efficient utilisation of data and be cost-effective. In the Netherlands for example, the government announced a call for selected university and hospital communities to develop technological and professional expertise in making data FAIR. The applicant communities had to cover 70 per cent of the investment costs.

> **The Committee recommends:**
>
> *Schemes should be established for encouraging collaboration on infrastructure and transfer of expertise between relevant communities in the form of networks, meeting places and/or centres.*

10.4 Investment in national data infrastructures

We use the term national data infrastructure in this context about infrastructure that is only affiliated to one institution or one community, but that all relevant communities in Norway can use. This type of data infrastructure is funded either directly or through competition-based funding. In

both cases, it is important that the investments are seen in conjunction with existing data infrastructures, both at the national and international level.

The Research Council is responsible, through the INFRASTRUCTURE initiative, for much of the investment in national data infrastructure. The initiative covers the establishment of new data infrastructure or upgrades of existing infrastructure. Many of the domain-specific data infrastructures are in an implementation and operation phase, and fall, strictly speaking, outside the initiative. A common denominator for many of the infrastructures is that they are a member of ESFRI – the European Strategy Forum on Research Infrastructures, where they attend to national commitments on behalf of Norway and the Research Council. As mentioned, the data infrastructures are knowledge-intensive and require regular updates in technical, data-oriented and scientific areas. Stable and reliable funding is needed to ensure that the data infrastructures can coordinate and deliver services at a high national level while also meeting the commitments of their ESFRI membership. The weakness of competition-based funding here is the time limit.

The evaluation of the INFRASTRUCTURE initiative concludes that it has had a structuring effect and has improved cooperation and synergies between actors. It has also helped to strengthen international collaboration and clarify the significance of FAIR data. However, a lot of work remains to achieve compatible solutions for data infrastructure. One example of this, which is very evident in the report on *Fremtidens miljødata* ("Future environmental data" – in Norwegian) (Menon, 2021), which highlights the need for better coordination of environmental data. There are many laborious solutions for finding and sharing data when it comes to environmental data, and different kinds of data are registered in different formats and in different ways. This is also pointed out in the FAIR study. It is important that work is coordinated centrally to prevent overlapping areas of responsibility and duplication of data infrastructure.

> **The Committee recommends:**
>
> *Any decision to establish and further develop national data infrastructures should be made at central government level either via the Research Council's INFRASTRUCTURE initiative or other mechanisms that ensure an assessment of the infrastructures' place in the national and international data infrastructure landscape.*

10.5   Why international participation is important for good coordination and division of responsibility

Norwegian participation in international collaboration on data infrastructure is important for Norwegian research because it gives researchers access to advanced infrastructure that Norway does not possess, access to better and bigger data, it facilitates collaboration between countries and it helps to develop and improve Norwegian infrastructures and pertaining expert groups. Most of the funding that takes place through international schemes is through Horizon Europe and DIGITAL. Participation in European networks and initiatives has given Norway the opportunity to coordinate national investments in data infrastructures with European investments in developing cross-sectoral infrastructure. Much of the development and implementation of EOSC has been carried out through EU-funded EOSC projects in the Horizon 2020 framework programme, which then continue into the Horizon Europe work programme for research infrastructure, where EOSC is a separate destination.

A majority of the data infrastructures on the ESFRI Roadmap are distributed. This means that different countries contribute complementary sharing infrastructure (known as nodes) in a common infrastructure. It is possible to apply for funding for infrastructures that help to meet the goals of the

ESFRI Roadmap through the EU programmes. Examples of international infrastructures that receive funding from the Research Council and EU programmes include

- ELIXIR Norway is the Norwegian node of the pan-European ESFRI infrastructure ELIXIR. ELIXIR Norway provides e-infrastructure for analysis, management and sharing of molecular life science data. Norwegian e-infrastructure for Life Sciences (NeLS) for non-sensitive data build on Sigma2, while solutions for sensitive data are established in collaboration with USIT-TSD, SAFE and HUNT Cloud. ELIXIR Norway develops and utilises ontologies and metadata forms that are used across Europe. It also provides data processing and advisory services.
- CLARINO is the Norwegian node of the ESFRI infrastructure CLARIN ERIC (Common Language Resources and Technology Infrastructure). CLARINO builds historical and contemporary electronic language resources and links language and text databases. The infrastructure links Norwegian databases with foreign databases, making it easy for researchers to download data with open licences (Creative Commons). One node in CLARINO is CLARINO Bergen Centre, which includes a storage and curation resource.

At an operational level, there are a range of international organisations, initiatives and projects that develop standards, solutions and infrastructures for increased sharing and reuse of research data. ESFRI, EOSC, RDA, OpenAire, GO FAIR and CODATA are among the important actors here. In 2019, five ESFRI cluster projects were established to enable the different ESFRI projects and landmarks to connect to EOSC. Several ESFRI nodes are affiliated to the cluster projects, including ELIXIR Norway through EOSC-Life and CESSDA through SSHOC. Some Norwegian research infrastructures are thus already involved and important in the establishment and implementation of EOSC through their participation in international consortia and networks. Their contribution to EOSC helps to draw attention to and make their own infrastructure available and attractive among European research communities. By contributing FAIR data and research results, Norwegian researchers can also draw attention to their own research and create new opportunities for interdisciplinary synergies and collaboration.

> **The Committee recommends:**
>
> *Efforts should be made to facilitate participation in international data infrastructure collaboration to ensure efficient use of resources and further development of Norwegian nodes and pertaining expert groups.*

10.6 Responsibility for making public administration data available for research

Many ministries and directorates collect data for administrative purposes. This data may be very valuable to researchers, but access can be a laborious and time-consuming process. In some cases, this may be because sharing the data as it is, is not straightforward and requires manual processing in each case. It may also be that the data are not equipped with good enough metadata, or that they have formats that are not compatible with the data the researchers want to combine the public administration data with. In some cases, the data can only be handed over to people, and not directly from computer to computer. In other words, more FAIR sharing of public administration data for research is needed than is the case today.

The strategy One digital public sector (2019–2025) defines the common goal that all public institutions shall share their data where possible. Two of its focus areas are: *"The public sector shall collaborate better on digital services and streamline the use of resources through enhanced*

coordination across administrative levels and sectors, and systematically realise benefits from digitalisation" and *"Data shall be shared and reused in the public sector to a greater degree, and open data shall be published for innovation and value creation in the business sector".*

We have many examples of administrative bodies making data available for research in different ways, e.g. the National Library of Norway, the National Archives of Norway, the Norwegian Meteorological Institute, the Norwegian Institute of Public Health and Statistics Norway. Some of them have come a long way in making data FAIR, while for others, the process of making data available for research is more time-consuming and expensive. The distinction between data infrastructures for research data and data infrastructures for public administration data is not always clear, which makes it expedient to see them in context. Several of the infrastructures, which largely offer public administration data for use in research, also provide an archiving service for research data in the same infrastructure. This applies to several of the infrastructures for environmental data, for example the Norwegian Marine Data Center. Another example is microdata.no, which now provides an opportunity for researchers to enter their research data in the infrastructure to be able to analyse them together with Statistics Norway data.

Extra efforts may sometimes be required to make the data applicable to research. There may be doubt about who is responsible for funding the extra efforts required. It could be argued that the Ministry of Education and Research should have main responsibility for this, and a few infrastructures have also received funding through the Research Council INFRASTRUCTURE initiative. The Committee is of the opinion that the sector ministries also have a responsibility to secure financial and organisational frameworks that facilitate public administration data for research within their area of responsibility.

The Committee recommends:

The sector ministries must take responsibility for making relevant public administration data available for research.

10.7   Capacity and priorities through networks and collaboration

As described in Part 2, making data FAIR is expensive, and the Committee has discussed to what extent the ambition should be for all research data to be FAIR. It will not be possible in the foreseeable future to make all data FAIR due to financial, technological and knowledge restraints. Within the individual research projects and groups, the research communities themselves must decide which data to keep. At a more overarching level, decisions must be made on which data collections and data infrastructures are to be prioritised initially. Such decisions should be made by the research communities together with other actors with an interest in access to the data. In some contexts, we also see different disciplines basing their research on the same type of data, but that have done so independently of each other. New interdisciplinary collaborations may require more coordination to secure a common understanding. Interdisciplinary collaboration will also help to raise competence and learning across disciplines, which will be important going forward.

The need for knowledge that can help to elucidate factors across societal and natural phenomena requires collaboration, cross-sectoral governance and funding. Individual ministries are often interested in promoting the interests of their sector and thus limit the comprehensive approach that is often needed to explore complex issues. This does not enhance interdisciplinary collaboration or help to find new solutions to complex problems. Several different inter-ministerial groups have been appointed, however, tasked with improving coordination between the ministries. The Committee

considers the proposal made by Skate (the inter-agency body responsible for managing and coordinating e-administration services) to the Ministry of Local Government and Modernisation on cross-sector governance to be sensible. To succeed, Skate recommends, among other things, coordinating letters of allocation, common budget items and that one ministry should have coordination responsibility for developments covering several ministries.

> **The Committee recommends:**
>
> *Arenas must be established for prioritising which data collections and data infrastructures resources should be invested in making FAIR. This must be done together with the research communities, public administration, trade and industry and other relevant actors in society.*

PART 4

Proposed framework for funding of data infrastructure

The Committee has developed a proposed framework over the course of its work that may contribute to more sustainable funding for establishing and operating data infrastructures. This part of the report sets out some overarching recommendations before going on to describe the framework in more detail. The framework is intended as an instrument that government ministries, funding agencies and infrastructures can use to find suitable funding models. The Committee underlines that it is not intended as a standard, but more as possible solutions that can be combined and adapted in each case. We see no reason to make changes to infrastructures that already have a well-functioning funding model.

<div style="border: 1px solid; padding: 1em;">

The Committee's overarching recommendations

- The Committee recommends using the framework for long-term and sustainable funding presented below to finance the establishment, further development and operation of national data infrastructure.
- The institutions should finance the establishment, further development and operation of data infrastructure that they are required to offer their research communities (basic infrastructure) over their ordinary budgets.
- National data infrastructures that receive long-term funding must have clear goals relating to the services they are to delive, and be evaluated regularly to ensure that they deliver in line with the goals that have been set for them.
- Investments in national data infrastructure should be coordinated and seen in conjunction with Norwegian participation in international data infrastructures, both in the form of Norwegian nodes and membership of international infrastructure collaboration.

The Committee's general recommendation is to continue well-functioning organisation, governance and funding mechanisms that already exist.

</div>

11. Proposed framework for future investment in and funding of data infrastructures

In the recommended framework, the Committee takes the great variation in existing data infrastructures into account, for example which needs they meet, how well developed they are, whether they offer generic or domain-specific data services, and whether they only offer local services or are open to national users outside the host institution. Since the needs the infrastructures cover differ to such an extent, the Committee believes it would be unrealistic and inexpedient to establish a "super infrastructure" to cover all needs, but that solutions are developed that make it possible to share and use data across infrastructures that serve different areas. The infrastructures' function, organisational model, cost level, capacity, user volume etc. will also affect how the costs linked to operation can and should be covered. For this reason, the Committee will not propose <u>one</u> way of funding data infrastructures, but rather a framework comprising several components that can be combined in different ways.

It is important to point out that a great deal of resources already go to investments in and operation of infrastructures and data services. The current situation works well for some infrastructures. In other areas, the current funding model is not very sustainable, especially seen in light of the fact that further development is required to comply with the FAIR principles. One of the purposes of the Committee's recommendations is to ensure cost-effective use of resources without this being at the expense of different user needs or limiting researchers' access to good, relevant infrastructures and services. An important means of achieving more optimal use of resources is to coordinate investments in infrastructure at the national level to prevent *unnecessary* parallel infrastructures and/or services being established at the institutions. More consideration should also be given to whether Norwegian communities should participate in international infrastructure collaboration, either in combination with or as a replacement for investment in national infrastructure.

Establishing, developing and operating data infrastructures is expensive if they are to meet knowledge needs and also utilise technological opportunities. Data infrastructures must constantly develop to meet society's and the research communities' needs. The requirement set out in the FAIR principles that data must be machine-readable increases the complexity of the infrastructures, for example. Regular investments and upgrades are therefore also needed, including in established infrastructures.

Long-term funding of a data infrastructure will ensure reliable operation, so that basic functions, including necessary competence can be maintained. Such guaranteed funding must not however produce a monopoly situation that does not take service quality and ROI (return of investment) into account. Regular evaluations must therefore take place, and the outcome must provide a basis for changing the course of organisation, funding and tasks. Critical user evaluations based on quality and costs are important for any data infrastructure, and the institutions must be able to select which services they want to fund and use. Requirements are also made in relation to annual reporting and formal contact with the user communities, which means real participation in further development.

User charges are recommended in some parts of the model, either as a one-off fee or subscription. For user charges to work, it is important that the research communities are given an opportunity to budget based on real cost estimates for data management when planning a research project. It is therefore important that the data infrastructures have price lists for the different services they offer that are included in the funding model. Preparing such price estimates can be difficult because it is impossible to predict future interest in and use of the infrastructure and pertaining services, but it is nonetheless absolutely necessary.

Proposed sustainable framework for funding of data infrastructures

Establishing, further developing and operating data infrastructures can include many different cost elements, for example procuring and upgrading physical infrastructure, operation of physical infrastructure, development and operation of digital services, user support, development of relevant competence for users and providers, continuous necessary upgrades and, if relevant, further development of the infrastructures. A sustainable funding model for such infrastructures must address the fact that these cost elements must be covered.

Breakdown into three service categories
Research data will typically undergo a dynamic life cycle during a research project, from planning and collection, via active use during analyses and processing, to archiving and, if relevant, sharing. Various infrastructure and service needs will arise during this process. The Committee's recommendations are based, in the following, on a rough overall breakdown into three categories of different types of tools, services or functionality (hereinafter "service") offered by data infrastructures during this life cycle. Which category the service falls under will affect what type of funding is expedient and realistic.

One single infrastructure may offer services in several categories, and the description of these categories is not exhaustive, and will not cover all types of functionality and versions of services affiliated to data infrastructures.

1   Basic, generic e-infrastructure that is a precondition for data-driven research. This will include physical infrastructure (for example high-speed networks and data storage), HPC and generic (interdisciplinary) software for analysing and processing big data.

2   Tools and services linked to active use, sharing and reuse of data. This may concern general and subject-specific infrastructures, but will largely concern subject-specific and discipline-adapted tools and services.

    2.1 Services that make data available for reuse in research. This includes publication of data, databases, data catalogues for searches, platforms and services for disclosure/access to data etc.

    2.2 Services and tools for active use, management, analysis and processing of data in research projects. This includes interaction platforms, metadata tools, ontologies and annotations, data management systems and electronic lab journals, programmes and services (pipelines/workflows) for reproducible analysis, visualisation and processing.

3   Services for long-term preservation and long-term management of data. This is often done in the context of a repository, which often offers many different services, but under this categorisation, the long-term management itself is a separate, basic service that is distinct from making data available (category 2.1). Infrastructure that offers such long-term management can be general, interdisciplinary or subject-specific.

Successful data-driven and sharing-based research is also dependent on good data services and systems that ensure, among other things, traceability and accreditation (PID issue, CRIS systems, citation tools, licences etc.) that are well integrated in the data infrastructure used. Many of these services are covered

through national and international common services combined with expertise and assistance in the support systems at the institutions, and are therefore not relevant in relation to the Committee's recommendations on investments in infrastructure.

Three main forms of funding

Infrastructures that form the basis for the services described above can be funded in a number of ways. The Committee has divided the funding into three main forms, which we will use in the recommendations below. In many contexts, funding will be a combination of these types.

1 **Long-term funding** – stable and predictable funding. The funding may, for example, be in the form of basic funding from one relevant ministry or from several ministries using a distribution key based on which administrative areas the infrastructure delivers research and management services to. Such funding may also be in the form of specific management assignments or through binding membership or partnership with institutions that use the infrastructure.
2 **Competition-based funding** – funding as a result of competition between several actors assessed on the basis of given criteria, for example the Research Council's infrastructure calls or the EU framework programmes.
3 **User funding or user charge** – funding that comes from those who use the services. This could be in the form of a subscription, a one-off fee or consumer funding, or a combination of these.

The Committee's recommendations

Based on the previous study, discussions in the Committee, input the Committee has received, and the findings presented in the [FAIR study report](), the Committee has made the following main conclusions on funding of different data infrastructures that are important for achieving FAIR research data:

1 **Funding of basic, generic e-infrastructure must be increased and be predictable in the years to come to meet the needs for this kind of infrastructure across the entire research sector.** Several sources, including the Research Council and the recent [evaluation]() of the Research Council's INFRASTRUCTURE initiative, conclude that more funding is needed for this type of infrastructure and that it should not be funded through competition-based funding. These are established infrastructures that provide services that are needed now and in the future.

    a) **A larger share of the funding of basic, generic infrastructure should come from long-term funding**. It is unfortunate that a large proportion of the funding of such infrastructure is competition-based. This creates uncertainty and a risk of losing important expertise, as well as a real danger of important research and administration tasks not being carried out. This type of infrastructure serves all discipline areas and projects in the university and university college sector, health sector and institute sector. The need will increase as research in all disciplines becomes more data-intensive and data-driven. Meeting the national guidelines for long-term storage of research and management tasks will require long-term and reliable funding.

    b) **Generic e-infrastructure services, in particular those that offer storage resources during a research project can be funded in full or in part by users.** User charge funding must then be included as part of the research projects' budgets, both in projects that receive external funding and in projects the institutions pay for themselves. It could be advantageous for large institutions that are involved in many projects generating large amounts of data to pay through institutional contributions to the infrastructures. This will reduce administration and ensure more stable and predictable funding of the national e-infrastructure that offers such services.

2 **It is recommended that data infrastructures for active use, sharing and reuse of research data are funded as a combination of long-term and competition-based funding and user charges.**

    a) **If the infrastructure covers a national need, funding parts of its operation through long-term funding can be considered**. Most of these infrastructures currently receive competition-based funding. As in point 1a, this will create uncertainty and a subsequent risk of losing important expertise. This is particularly unfortunate for some, for example those that are part of an international infrastructure in which Norway has undertaken to participate. Data infrastructures are subject to specific requirements relating to the services they must deliver in order to receive basic funding. These requirements may concern making data FAIR, making data complementary to other data, that data must generate added value for the discipline beyond existing services financed through basic funding and approved international data infrastructures, and that technical requirements must be made of the service. This means that we can build on existing services, and thus safeguard domain and institutional heterogeneity at the same time as ensuring that every institution or individual project does not create its own data infrastructures and services. The Committee does not reflect on who should assess whether a service meets these requirements, but would like to point out that this is comparable to the job the Research Council does of assessing whether research institutes meet the criteria for being awarded basic funding, and also, in part, assessing to what extent a research infrastructure meets the requirement of being national. Several international certification standards, for example Core Trust Seal, GoFair and EOSC, may also be relevant in this context.

    b) **Establishment of data infrastructures or extensive expansion or further development of a data infrastructure's services for sharing and reuse of research data are made subject to competition.** This can currently be done through the Research Council's INFRASTRUCTURE initiative or through other competition arenas. The advantage of awarding funding through competitions is that the infrastructures are assessed based on clear requirements set out in the call. This will identify and promote infrastructures that address national needs within specific disciplines or technical solutions adapted to the needs of research, the public administration, trade and industry, and the voluntary sector. Gaining a qualified assessment of the added value of developing the range of services is important, and competitions that include strategic and discipline-based assessments are therefore suitable.
To ensure permanent operation of a newly established infrastructure, funding can come from different sources depending on the need the infrastructure intends to meet. The model for future operation must be assessed before a decision is made on awarding funding to the new establishment. Established data infrastructures can, when well established, qualify for category 2a funding.

    c) **Services that are geared towards individual projects or are closely linked to physical infrastructures can be funded by user charges.** This may for example be in the form of adapted services that provide researchers with access to processed or collated data, which it is natural for a project to pay for. It could also be services linked to the submission and publication of research data in line with the FAIR principles. Such costs must be included as part of the project budget if it is externally funded. Institutions that make requirements of FAIR research data from their researchers should then also expect to enter this in their budgets, either by "charging" the costs to individual projects or through institutional contributions to relevant data infrastructures. It will be natural for the infrastructures, to the

extent that the forthcoming implementation of the Open Data Directive in Norwegian law allows, to include an overhead for the services to cover part of the operating costs through such a user charge.

3 **Funding of repositories for the long-term preservation and curation of valuable research data and public administration data of relevance to research must be a national responsibility and awarded long-term funding**, as is the case to a certain extent in the ALM sector today. Long-term preservation cannot be covered by individual projects with a limited budget period, and many types of data may have value beyond an individual research project. Predictable funding is also needed to secure long-term repository services for research, the business sector and public administration. This means that such repository services should mainly have long-term funding. Whether this should come from one or more sources must be assessed. Relevant sources of funding are all ministries with a sector responsibility for research. The Ministry of Education and Research, which has the main responsibility for research and education (including infrastructure), must take responsibility for coordinating the ministries in a manner that supports the coordination of long-term funding. This can also be combined with a requirement that all data-generating institutions contribute funding to the infrastructure themselves.

An example

We have used Statistics Norway's function as a data provider for researchers to demonstrate how a model like this will work in practice. We have categorised different Statistics Norway services in that connection in the table below, and described how the different services can be funded based on the proposed model.

Service category/activity	Collection and management of statistics data	Making adapted data available for researchers in the conventional manner	Establishment and further development of microdata.no	Operation of microdata.no
Form of funding	Long-term funding	User funding, one-off fee	Competition-based funding	Partial user-funding planned

Competition arenas that secure development and modernisation of technology, services and competence

As well as investing in the infrastructures themselves (equipment and personnel) that will lead to more FAIR research data and relevant public administration data, the Committee also sees a need to increase investments in the development of new technology and new services to make the infrastructures relevant and internationally competitive for future research needs. The funding of such development projects must be seen in close context with investments in the infrastructures. These projects often fall under established funding schemes because we do not have good enough mechanisms for identifying, assessing and quality assuring them. Defining projects under established categories can be difficult because they are often found in the intersection between research, methodology development, service development and infrastructure establishment. A lot of uncertainty and risk is often associated with such projects because they seek to find solutions to partly unknown, future needs and they may be experimental in nature. As such, they often do poorly in competition with other, less risky projects. Examples of such projects could include the development of new computational technologies, such as neuromorphic computing, analysis methods for large, complex amounts of data based on machine learning and artificial intelligence, network services in line, for example, with the development of 5G/6G networks or new FAIR repository solutions. The Committee is therefore of the opinion that established competition arenas should be extended to

include development projects, or that separate arenas are established with criteria that enable risky projects to succeed in the competition.