"… the future of science: a global data commons,
a virtual science library spanning the globe."

RDA Europe [1: p.5][1]

SCIENTIFIC PLAN: THE SWEDISH NATIONAL DATA SERVICE
For Sharing and Reusing Research Data

Sweden needs a well-functioning infrastructure that support the sharing and reuse of research data. Policy makers and funders of research, both national and international, require that the research output they are funding is made public. Advocates of open access to research data point to advantages in terms of research transparency, impact, and cumulative science. Other benefits include sustainability, efficient use of public funds, and more equitable access to high-quality data. Nevertheless, there are significant barriers to sharing data. Technical solutions and community practices, as well as legal requirements and ethical standards, all pose different challenges to researchers who try to share their data in a way that makes these data findable, accessible, interoperable, and reusable (the FAIR principles) [3].

Since 2016, the Swedish National Data Service (SND)[2] has worked in collaboration with Swedish universities and research organisations to build the foundation for a nation-wide system that can support researchers in providing access to their data. By prompting the establishment of local research data support functions (in this application referred to as DAUs [Data Access Units]), by providing their staff with training, and by introducing them into the *SND Network*, SND established a framework by which all researchers could have access to qualified support in research data management and sharing. In the next funding period (2023–26), SND will take the next steps to towards a well-functioning and qualitative infrastructure by ensuring that the Swedish research community can

- describe and preserve research data, including sensitive data;
- share research data with Swedish and international researchers; and
- find openly available research data to reuse.

The responsibility for handling and storing research data in accordance with Swedish law lies with the university where the data were created. The universities should also ensure that data align with the FAIR principles and that, by 2026, all publicly funded research data are made accessible as open as possible but as closed as necessary [4]. To achieve this, all universities need to provide their researchers with support and training in research data management, as well as with adequate data storage for the types of data that they produce. SND's role as a national infrastructure for supporting open research data is to be a knowledge hub for data management at the centre of a system for disseminating and exchanging know-how, expertise, and experience. The infrastructure also provides research data producers, research data users, and research data support functions in Sweden with technical tools and services that facilitate describing, sharing, and reusing FAIR data.

SND is an infrastructure that supports all kinds of research and for all researchers in Sweden. Unlike the type of research infrastructures implied in the instructions for this application, SND does not have a distinct focus on formal research groups, or research fields, but, rather, caters to all types of researchers and scientific fields. In order to give an accurate description of SND's goals and operations, we occasionally need to deviate from the instructions from the Swedish Research Council for this call.

[1] For references, see *Appendix E: Key References*.
[2] Link to the SND website: https://snd.gu.se/en.

1. Motivation and Overview

1.1. An Infrastructure for Digital Research Data

In the 2020/21 research and innovation bill, the Swedish government emphasised that results from all publicly funded research, publications as well as data, shall be made freely accessible from 2026. Open access shall be the norm, except when there are reasons to restrict data access, such as protecting privacy, national security, or intellectual property. The bill also indicates that universities should harmonise their principles for open access to publications and data as far as possible, and ensure that data align with the FAIR principles [4]. It could be impossible to achieve these ambitions efficiently without a national infrastructure for research data such as SND.

Swedish research initiatives and infrastructures have made some progress towards FAIR and open data in recent years. However, much remains to be done before Sweden reaches the national and European goal of making open access to data (created in whole or in part by public funds) the norm. Swedish official reports have observed a need for increased national coordination of e-infrastructures[3] to the point where a new public agency for e-infrastructure is currently under inquiry [5]. Improved national coordination is a basic condition for being able to support the entire research process, from data collection to long-term preservation and access, and it is a necessity if Sweden wants to actively influence the European agenda on research data infrastructures [5–7]). As the most prominent Swedish organisation that promotes operational cooperation between universities' and research e-infrastructures' open-data efforts, SND is a key actor in making Sweden reach its open-access goals.

There is a clear and pressing need for all researchers to be able to practice sustainable, open science. Through open and transparent data sharing, research can become a truly cumulative knowledge-creating process. Scientific journals, public and private funding agencies, university policies, and legislators already require research data to be made accessible. In the future, researchers who do not comply with data access requirements may face difficulties funding projects or publishing results. SND aims to be a national infrastructure that enables researchers in Sweden to be forerunners in the dynamic and quickly developing landscape of FAIR data and open science. To achieve this, the infrastructure will provide coordinated and secure structures for describing, depositing, sharing, and finding high-quality, open research data that adhere to the FAIR principles.

The aim of the SND infrastructure translates into three overarching goals outlining the infrastructure's position within the Swedish research community:

Facilitate the sharing of research data. SND will further Swedish research by providing essential, first-rate support and systems for describing and sharing data in a way that is simple, secure, and trustworthy – and FAIR. Thus, compliance with journal, funder, university, and legislator requirements (current and future) will not place an unreasonable burden on the researcher. However, systems and processes will protect sensitive data and the privacy of research subjects, and data integrity will be ensured through certified curation and preservation workflows.

Provide data visibility and usage metrics. The infrastructure will be a core service for researchers in making data visible, nationally and internationally. Increasing the visibility of data is a key factor in furthering the impact of Swedish research. Tools for providing metrics on the sharing, reuse, and citation of datasets will be offered. In the future, the reuse of data might become an important factor in determining scientific impact, in the same way that for example journal impact factor and h-index are to certain disciplines today.

Facilitate trustworthy data access. The infrastructure will provide a national search portal for finding high-quality, well-documented data for use in research projects. Data provided in

[3] Although SND is not formally an e-infrastructure, it is occasionally referred to as such (e.g. [5]).

the portal could be used, for instance, for new analysis, validation of previous studies, preparation of new studies, and creation of aggregated datasets. The portal will provide information on the conditions for access, use, and reuse of the accessible data. Thus, it will be the main discovery tool for accessing Swedish research data.

The objective of SND is to be the driving force in: building a system of technical solutions, creating organisational networks, fostering fruitful collaborations, developing reasonable and accurate metrics, and providing highly-skilled local and national support. In order to achieve this, SND needs to further develop and maintain cutting-edge expertise on data management and curation, provide high-quality training to data professionals and researchers, and spearhead the development of methods and best practices as the research data landscape evolves.

1.2. Major Challenges for Society and the Research Community

Many of the challenges that our society faces today are global and require global research. To tackle major crises such as the COVID-19 pandemic or climate change, data from around the world need to be combined and shared, and therefore available for quick access. Funders are expected to increase demands for research data that can be reused in a way that contributes as much as possible to public good, therefore necessitating open access to data.

Not all research data are suited to such easy access, however. Large datasets, exceeding the technical limitations of web browsers or even FTP clients, is one example where restricted access is necessary. Some research centres even create such enormous datasets that moving them elsewhere for preservation and dissemination is inefficient; they must be disseminated from the point of creation. As each new generation of datasets tends to be larger than the previous, the future will place growing demands on data curation proficiency, dissemination resources, and the technical know-how required to provide access to these large datasets. The rapid changes of metadata standards and best practices in some fields may also require restricted access; in emerging or cutting-edge research areas, the development of standards is by necessity a constant process. The movement from static datasets to dynamic data (that expand and develop over time) adds additional challenges to open access. Finally, disseminating research data also involves considering limitations on when, with whom, or under what conditions data can be shared. Confidentiality provisions, contractual obligations, embargos, and/or sensitive personal information can all warrant restrictions.

In cases such as those exemplified above, curation, legal expertise, and technical systems are necessary in order to make correct decisions and to ensure that data are delivered safely and within limitations. Staying on top of all the changes and maintaining sufficient quality, in terms of documentation as well as file formats, is a highly resource-intensive activity, and requires both national and international collaboration and pooling of resources.

During the current funding period (2018–22), SND has created a coordinated, national infrastructure that can take on these various challenges. Training and knowledge exchange within the SND Network – currently DAUs at thirty-five universities and three public research organisations – will intensify during the next funding period. (How this work will develop is outlined in *Description of the Infrastructure and its Activities*.) In order to meet societal challenges more broadly, SND has initiated discussions with business organisations regarding participation in the SND Network. SND will also continue its active participation in the European Open Science Cloud (EOSC)[4] and its related projects. (See also *Appendix C: Activities during the Current Funding Period.*) The challenges that society faces are best met with access to open, transparent research of highest quality; and the demands that this places on the research data community are best dealt with together.

[4] For details about EOSC, visit https://eosc-portal.eu/

1.3. LONG-TERM VALUE

Since its inception as SSD (Swedish Social Science Data Service) forty years ago, SND has been working to safeguard, share, and preserve data created by researchers at Swedish universities. At the core of the consortium is a vision of allowing researchers to share well-structured, well-documented, and easy-to-find research data, now and in the future. To realise this vision, SND will continue to develop and share its expertise, skills, and knowledge about data management and curation through various training activities for data professionals. Close collaboration between SND and the DAUs at Swedish research organisations will in the future provide all researchers in Sweden with local access to trustworthy research data repositories (for details, see submodule 4.3 in *Description of the Infrastructure and its Activities)*. National and international collaborations will intensify in order to accumulate technical knowledge and other know-how that will benefit research communities in Sweden. In this respect, SND will continue to serve as the national service provider for CESSDA ERIC[5] as well as being the national node for the Research Data Alliance (RDA).[6]

1.3.1. Relations to Other Infrastructures

SND is the only national infrastructure in Sweden with the ambition to support sharing and preservation of research data across all research domains. There are areas in which domain-specific data infrastructures exist, but they are comparatively few and often with a narrow scope (areas include astronomy, archaeology, high-energy physics, endangered languages, corpus linguistics, climate research, and others). SND sees collaborations with such specialised infrastructures as the natural way forward, both in terms of metadata harvesting and knowledge exchange. Internationally, SND has long-standing collaborations with other national research data repositories, mainly in Europe and with particularly close ties with the national repositories in the other Nordic countries (for details, see submodule 3.2 in *Description of the Infrastructure and its Activities*).

As the only national data access infrastructure working across all scientific domains and in collaboration with nearly all Swedish higher education institutions, SND is uniquely positioned, experienced, and prepared to undertake coordination in this area. Such national coordination would for instance include work with EOSC, where SND could collaborate with SUHF (Association of Swedish Higher Education Institutions), VR (Swedish Research Council), and other Swedish EOSC members.

1.3.2. An Infrastructure Serving Researchers in Sweden

The current SND infrastructure is already an important resource for many research efforts. The SND research data catalogue[7] enables disciplinary and interdisciplinary searches and contains high-quality data and metadata. In the catalogue, data are efficiently and securely disseminated, even when access is restricted; and data descriptions can be made visible in international data registries. SND's services and systems are continuously developed to permit metadata and data to adhere to the FAIR principles. The current infrastructure has developed from a centralised repository with restrictions to a new architecture with the ability to provide a repository solution that can host personal data as well as large-size data.[8] Through its domain specialists and research data advisors, as well as with its support to the local DAUs, SND helps researchers comply with institutional (e.g. university, legislator) and professional (e.g. journals, funding agencies) requirements on data sharing and preservation.

---

[5] For more information about CESSDA ERIC, visit https://www.cessda.eu/About. For details about SND's role as CESSDA service provider, see submodule 4.4 in *Description of the Infrastructure and its Activities.*

[6] For more information about RDA, visit https://www.rd-alliance.org/

[7] The SND catalogue can be found at: https://snd.gu.se/en/catalogue

[8] For more details on sensitive data and data storage, see *Description of the Infrastructure and its Activities*.

Private and public research funders and scientific journals require accessible data, and many researchers thus need reliable and simple solutions. The SND infrastructure is listed among recommended repositories at e.g. PLOS One[9] and Ubiquity Press,[10] providing Swedish researchers with a national option for complying with journal requirements for data availability.

As the research data landscape evolves, SND's systems will continue to adapt to the specific needs of various disciplines in terms of metadata profiles, FAIRness, and tools, in preparation for a future where SND is a researcher's obvious first choice for describing and sharing important research data. SND is a Trusted Digital Repository, certified with the CoreTrustSeal,[11] a guarantee that data are documented, curated, and stored in a manner that safeguards their value as a reusable resource for future researchers. In accordance with the current model for research data support from DAUs at universities and research organisations, SND will continue to provide assistance and expertise to enable every DAU to become certified as a Trusted Digital Repository. More importantly, SND will continue to develop systems that make data documentation as quick and easy as possible. A significant step in that development will be strategies, workflows, and functionality to augment entry-level data descriptions to greater degrees of FAIRness (entry-level descriptions were introduced during the current funding period; see also submodule 4.2 in *Description of the Infrastructure and its Activities*).

In preparation for the national goal of open access to research data by 2026, SND will develop a new national research data portal, *researchdata.se/forskningsdata.se.* The goal is to make it easier to find, access, share, and reuse data. The portal will be developed to be an obvious first port of call for any researcher or organisation looking for Swedish research data, data-management-related information, or gateways to national and international resources (for details, see submodule 4.1 in *Description of the Infrastructure and its Activities*).

As researchers' data become more visible through the new national research data portal, SND will develop analysis tools that can provide researchers with metrics for usage and scientific impact. Examples include data citation and data download metrics. The national portal will also allow us to collect nation-wide statistics about data usage, for instance number of citations and number of orders and downloads.

1.3.3. An Infrastructure of National Interest

When researchers have to comply with increasing requirements on data access and preservation, compliance must be easy. Most researchers may not need to describe data very often, and encountering a new interface each time would serve as a disincentive to doing a good job. By building and maintaining a national system for describing and providing access to research data, SND offers the same interface to all researchers at Swedish universities, regardless of where they are currently employed. Researchers can create data descriptions using their existing university login credentials. The same metadata schemas and tools are accessible to everyone. Describing data should be as easy as possible.

Simplicity and equal conditions go beyond describing and sharing research data. Through the SND Network, researchers will also encounter similar research data support, regardless of where they work. The network will ensure that all DAUs have access to the same kind of support from SND, that expertise is disseminated across all network members, and that collaborations drive improvements in data management and curation on a national rather than local level. SND's networks of university lawyers, archivists, and IT specialists strive to offer similar assistance with regard to data-related issues, and the consortium will develop guidelines and best practices to provide uniform advice to all researchers in Sweden.

---

[9] The PLOS One list can be found at: https://journals.plos.org/plosone/s/recommended-repositories.
[10] The link to Ubiquity Press is https://www.ubiquitypress.com/site/. See for example Journal of Open Archaeology Data; Journal of Open Humanities Data; Journal of Open Psychology Data.
[11] More about the CoreTrustSeal on the following link: https://www.coretrustseal.org/

As a national infrastructure, SND develops tools and processes meant to be used nationally. Through collaborations between consortium partners, SND works to identify what researchers in particular domains or research fields need in regards to providing data access. Such initiatives will then result in tools, resources, and services (e.g. support, training, review assistance) based on those needs. Because SND is a multi-disciplinary national infrastructure, the benefits of improvements made by addressing the needs of a certain group become available to all researchers in Sweden, or even internationally. Although much research is international in nature, and even more will be so in the future, there are some domains in which Swedish data are of particular interest. They may concern specifically Swedish issues, make use of particular Swedish advantages, or be regulated by Swedish legislation.

Without national coordination, the implementation of free access to research data risks resulting in a fragmented system that is inequitable, inefficient, unequal, and lacks interoperability. It would use up valuable resources when multiple players solve the same problem in different ways. It would also create unnecessary work, and ultimately lead to less data reuse and be an impediment to world-class research. The SND consortium is uniquely positioned to provide a national infrastructure of benefit to researchers in Sweden and research in general.

1.4. Current and Future Users of the Infrastructure

Since 1981, SND and its predecessor SSD have accumulated considerable expertise in making data accessible for various users in accordance with the OAIS reference model.[12] SND provides access to social science data since nearly four decades, works with data from the humanities and health sciences since 2008, and hosts Environmental Climate Data Sweden (ECDS) since 2016. Starting at the beginning of the current funding period, SND expanded to provide support for an increasing amount of research domains by shifting to an organisational model in which expertise is distributed over a consortium of major Swedish universities (for details, see *Appendix C: Activities during the Current Funding Period*). The long-term objective is to establish a nation-wide system that allows researchers to share well-structured, well-documented, and easy-to-find research data from all scientific disciplines.

Currently, SND has three main categories of users: **research data producers** (mainly researchers), **research data users** (researchers, businesses, public authorities, etc.), and **research data support** (local DAU staff and other data professionals). The next funding period will see an increase in activities towards the first two categories and a clearer separation of tools, services, and resources aimed at the respective categories. This separation is prompted by an expected increase in the number of users, a greater number of collaborative initiatives to meet the needs of these users, and a rising complexity of research data issues (for details, see section 3.3 in *Description of the Infrastructure and its Activities*).

The SND catalogue with research data descriptions will provide the core of the future *researchdata.se* portal (for details, see submodule 4.1 in *Description of the Infrastructure and its Activities*). Over the first three years of the current funding period, 325 data descriptions were added to the catalogue (73 in 2018, 86 in 2019, and 166 in 2020). The catalogue currently (12 February 2021) contains 1,656 data descriptions from a variety of research areas. A large share of the studies are available for direct download. The total number of downloads from the SND website in 2020 was 26,447.[13] The series with the most downloads in 2020 were International Social Survey Programme (ISSP), Institutional Trust, the Swedish part of the European Social Survey (ESS), and Arctic Ocean 2016.

---

[12] The OAIS reference model includes: Ingest, Storage, Data Management, Administration, Preservation Planning and Access. https://public.ccsds.org/pubs/650x0m2.pdf.

[13] Downloads from catalogue posts with accessibility level **"**Access to data through SND – Data are freely accessible". Other posts may have documentation available to download (even if the data have to be ordered). Organisation is not registered for users that download directly from the catalogue.

Material that is not available for direct download from the SND website can be ordered upon request, followed by a brief assessment before delivery. In 2020, SND received 438 orders, 406 from academic users and 32 from non-academics.[14] These orders comprised 1,682 datasets.[15] Academics from the University of Gothenburg dominate, followed by users at Uppsala University, Stockholm University, Lund University and Mid Sweden University. Due to legal restrictions, SND has limited statistics concerning orders from non-academic users. These users represent a variety of organisations, such as private companies, media, schools, local and regional public administrations, government authorities, political parties, NGOs, research organisations, labour market organisations, think tanks, libraries, and archives. Studies that are part of a series were among the most popular. The series with the most orders in 2020 were the National Society Opinion Media (SOM) Survey and the Swedish National Election Studies, followed by SVT Exit Poll Survey and Swedish Electoral Data.

Several research groups with internationally renowned scholars use the current SND services to provide access to their primary data and to use secondary data for new analyses. (The publications are listed in *Appendix E: Key References*.) These include the top ordered and downloaded datasets from: several Swedish *electoral studies*[16] [8–10]; the *SOM surveys* [9, 11–12]; the *Media Barometer* [11, 13]; *Census of the population* [14]; *Arctic Ocean 2016* [15–16]; *Oden Southern Ocean 2007/2008* [17]; *Reflection seismic study of the Siljan Ring impact structure: Mora* [18]; *Brexit Blog Corpus* [19]; *NordChild* [20–21]; *Swedish Contextual Database for The Swedish Generations and Gender Survey and The International Generations and Gender Programme* [22–23]; *Noise exposure files for a polysomnographic study of ground-borne noise from railway tunnels and sleep* [24]; *Observations of Noctilucent Clouds from Denmark during 2011* [25–26], and the *DREAM Dataset: Behavioural data from robot enhanced therapies for children with autism spectrum disorder* [27].

1.5. Building National Data Management Expertise

A crucial part of the SND infrastructure will be the knowledge exchange that it facilitates. In the rapidly-changing landscape of open data, there must be a way to quickly disseminate new skills, experiences, and expertise. Even though SND implements strategic measures to keep abreast of new developments in research data management and curation, the task is too big for SND staff alone. Instead, SND will be organised to enable effective knowledge exchange between the various partners in the SND networks. Trainings will be arranged for network members. There will also be seminars and workshops to allow members to exchange new knowledge and data management experience with others, including SND staff.

Joint learning resources will be developed within the various networks. Professional development courses will be maintained, and SND will expand collaboration with partners in developing courses for doctoral programmes and for data professionals on a national and local level. The focus for these resources and training initiatives will shift during the course of the funding period, with a relatively smaller focus on DAU staff as the DAUs reach maturity.

The Open Data Flagships and other targeted Open Data initiatives that will run at the consortium universities in the next funding period will also help contribute to a broader national knowledge about data management and related skills (more about this in section 2.1 in *Description of the Infrastructure and its Activities*). The Flagships in particular will include activities relevant to disseminating expertise in managing research data efficiently, securely, legally, and FAIRly among researchers in Sweden.

SND's training and knowledge exchange strategy will contribute to data professionals and researchers at Swedish universities and research organisations having gained skills and

[14] SND currently does not collect statistics on users' academic disciplines.
[15] SND does not collect information on the gender of its users.
[16] Including Swedish National Election Studies; European Parliament Election Studies; SVT Exit Poll Survey; Swedish Electoral Data.

knowledge in how to manage their data well. As a result, they will be able to share and preserve well-documented research data that can be used by future researchers.

2. An Infrastructure to Promote Research, Innovation and Public Good

2.1. Benefits for Swedish Innovation and Efficient Use of Public Resources

Research output is of interest to private enterprises and the public sector, as well as to the academy. Access to publications and data from the universities have proved to be an expensive, time-consuming, and piecemeal affair for actors outside of academia. Similar to how open access to publications offers non-academic actors much better access to research findings, the SND infrastructure will offer companies and public agencies vastly improved access to the related research data.

SND provides a system through which digital research data can be found and accessed by private and public actors. Well-documented data include clear information about how data can be reused and how to gain access to the data if they cannot be downloaded immediately. Because research data are collected from all disciplines, it is not possible to offer a succinct description of which particular innovations or innovators can benefit from SND, nor what social or public utility the data can create.

A notable contribution to society of the SND infrastructure is the advancement of FAIR data. FAIRness promotes efficient use of public resources by ensuring that research data are documented and preserved for long-term, future use, thus maximising the value of research funding. According to a report published by the EU Directorate-General for Research and Innovation [28], a comprehensive cost-benefit analysis estimates the total annual loss of not having FAIR research data to more than ten billion euros in Europe alone, or about four fifths of the annual Horizon 2020 budget. The report associates similar costs to not having open access to data. From this perspective, the Swedish goal of FAIR and open research data by 2026 has a distinct economic dimension that goes beyond the economies of scale associated with a national research infrastructure.

2.2. Contributing to the Global Sustainable Development Goals

The United Nations' Global Sustainable Development Goals 2030[17] (SDGs) offer a roadmap for a better tomorrow for humanity, and research is key to getting us there. Through research, challenges and conflicting objectives can be identified, solutions found, dialogue between different societal stakeholders initiated, and the goals themselves analysed. The SND infrastructure does not contribute to specific goals, targets, or activities as much as to the entire system of research – in a wide range of areas – that help us progress in the right direction.

A research infrastructure that provides researchers with the possibility to make their data FAIR is ultimately essential if we are to efficiently meet the SDGs. The challenges encompassed by the seventeen goals are global in nature, requiring an equally open research effort. Some research fields are already leading the way in terms of openness and data sharing, such as the environmental and climate research required to combat climate change and its effects (Goal 13). Through the SND infrastructure, other areas can now follow suit.

An example of how SND has contributed specifically is the collaboration with Environment for Development (EfD), a global network of research centres solving environmental and developmental challenges. Within the framework of the collaboration, EfD have gained expertise and developed processes necessary for sharing well-documented data. This has included training data managers and curators at Policy Studies Institute in Addis Ababa, to improve data preservation and sharing within the network. Similar collaborations that contribute to research that further the work towards the SDGs will occur over the next decade.

[17] For more information about the UN Sustainable Development Goals, see https://www.un.org/sustainabledevelopment/