



Open access to research data in practice

The Swedish National Data service, **SND**

David Rayner & Lisa Isaksson



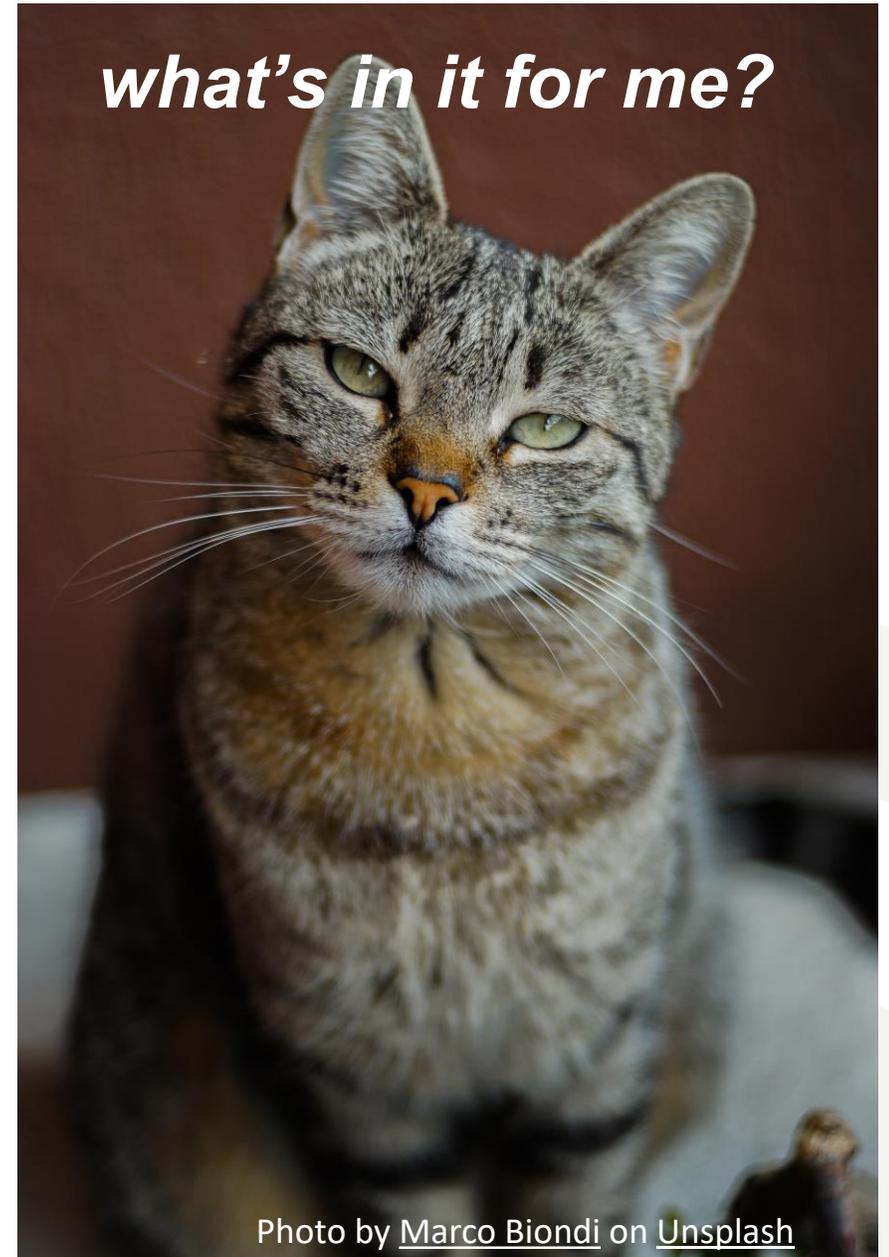


Presentation overview

- Why make research data accessible?
- Introducing **SND**.
- “Open access to research data” in practice:
 - As open as possible
 - The FAIR principles
- Data management plans – systematic data handling.
- What does **SND** provide?

Why share research data?

- Avoid disasters.
- Increased impact and visibility.
- Be seen as more trustworthy!
- Improve article citations!
- Get credit.
- For your future self!
- Do something good for the world.



Who benefits from open access to research data?



The Academic Community

- Basis for new studies
- Feasibility studies, project proposals
- Initialization, calibration, verification
- Teaching
- Promotes interdisciplinary collaboration

Democratic society

- Publicly funded research
- Avoid useless duplication
- Research transparency and reproducibility



Swedish National Data Service, **SND**

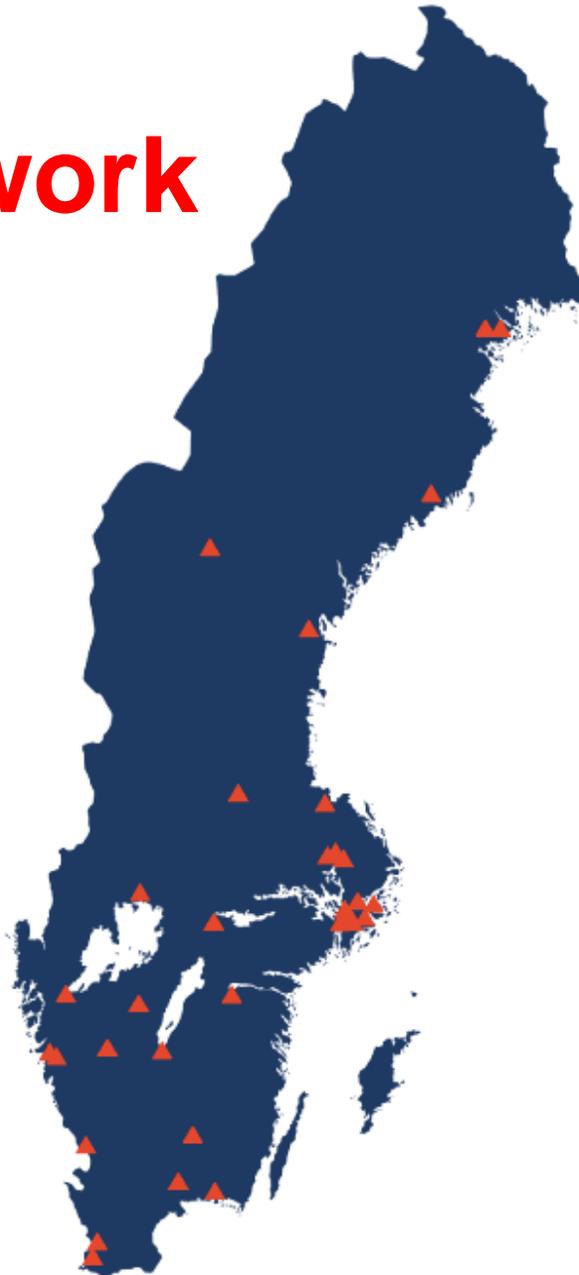
A national research infrastructure for making research data accessible.

- **SND** provides research data management support
- Support via the **SND network**
 - Universities
 - Research organizations



The **SND Network**

- Around 40 higher education institutions and research organizations.
- Governed by a consortium of 9 universities.



Luleå University of Technology

Swedish Polar Research Secretariat

Umeå University

Mid Sweden University

University of Gävle

Uppsala University

National Veterinary Institute

Swedish University of Agricultural Sciences

Mälardalen University

Stockholm University

Karolinska Institutet

KTH Royal Institute of Technology

Stockholm School of Economics

Sophiahemmet University

Swedish Defence University

Swedish Geotechnical Institute

The Swedish School of Sport
and Health Sciences

University College Stockholm

Södertörn University

Linköping University

Jönköping University

Linnaeus University

Dalarna University

Örebro University

Karlstad University

University of Skövde

University West

University of Borås

University of Gothenburg

RISE Research Institutes of
Sweden

**Chalmers University of
Technology**

Halmstad University

Blekinge Institute of
Technology

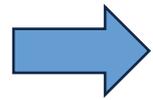
Lund University

Malmö University

Kristianstad University

Sweden

Why does **SND** exist?



Open Access is a Swedish Government policy in the transition to an open science system!

“ Research data that is produced by publicly funded research should be made accessible according to the principle: "as open as possible, as restricted as necessary". ”

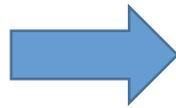


[The Swedish Research Council – The way towards open access to research data](#)

What is “open access” to research data?

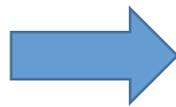
Two (overlapping) sets of principles:

Who has access?

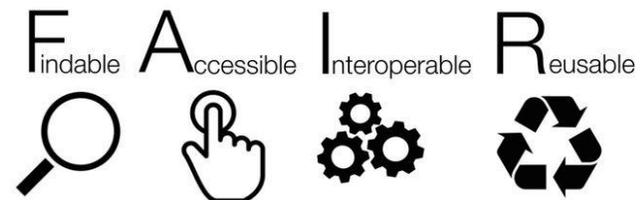


The “as open as possible, as restricted as necessary” principle.

How are the data made accessible?



The **FAIR** principles.





“As open as possible, as restricted as necessary”

Is it possible to make a dataset **freely** (openly) accessible?

Reason for restrictions:

- Sensitive data
 - protected species
 - classified information
- Personal data
 - whether sensitive or not
- Size/complexity



Photo by Noun Project from Noun Project

What is **personal** information?

“ Information that directly or indirectly refers to a living, identified or identifiable natural person.”

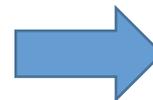
- “Indirectly”:
 - recorded voice
 - photograph where person is identifiable
- Combination of information → risk for “re-identification”:
 - Age + occupation + town



Pseudonymized data are still personal data!

- What is “pseudonymized data”?
 - no direct identifiers
 - tagged with codes or serial numbers matching a *separately stored* “code key”.
- Pseudonymized data are personal data!
- Even if you as a researcher cannot access the code key!
- Cannot be shared freely!

Alice	2h
Bob	3h
Tom	7h



Created by Kreev Studio
from Noun Project

001	Alice
002	Bob
003	Tom



001	2h
002	3h
003	7h

Anonymizing (de-identifying) data

Anonymization: Removing everything that identifies people, directly and indirectly, so that it is **no longer possible to link the data to an individual.**

- Suppression/masking – remove!
- Tokenization
 - Replacing identifiers consistently.
- Generalization – reduce granularity
 - Categorizing age into age groups.
 - Recoding rare categories as “other”
 - Replacing address with the name of town/municipality/region



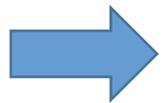


Anonymizing (de-identifying) data

- Interpretation differs within EU

! "Anonymized" dataset might not be considered anonymous if original dataset retained!

! Archiving or replication regulations may mean you *must* retain original dataset!

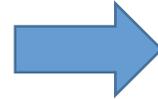


If you don't need to know identities, collect data anonymously (where possible)!



Is it possible to make a dataset **freely accessible**?

- Can you anonymize the data without significant information loss?
- Did you specifically obtain informed consent to share the data *for unrestricted purposes*?



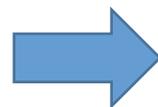
Freely
accessible

Is it possible to make a dataset **accessible for research purposes?**

- Can personal data be *re-used* for research?

→ Yes! GDPR [Article 89\(1\)](#)

- research “not considered to be incompatible” with the initial purposes.

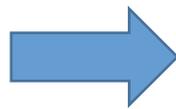


Restricted
access

What is “open access” to research data?

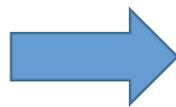
Two (overlapping) sets of principles:

Who has access?

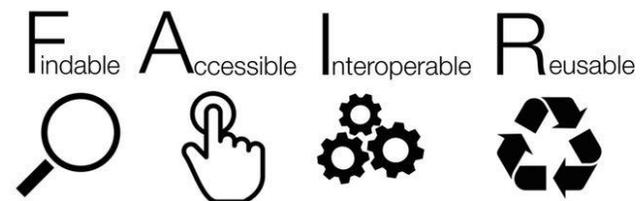


The “as open as possible, as restricted as necessary” principle.

How are the data made accessible?



The **FAIR** principles.



FAIR principles

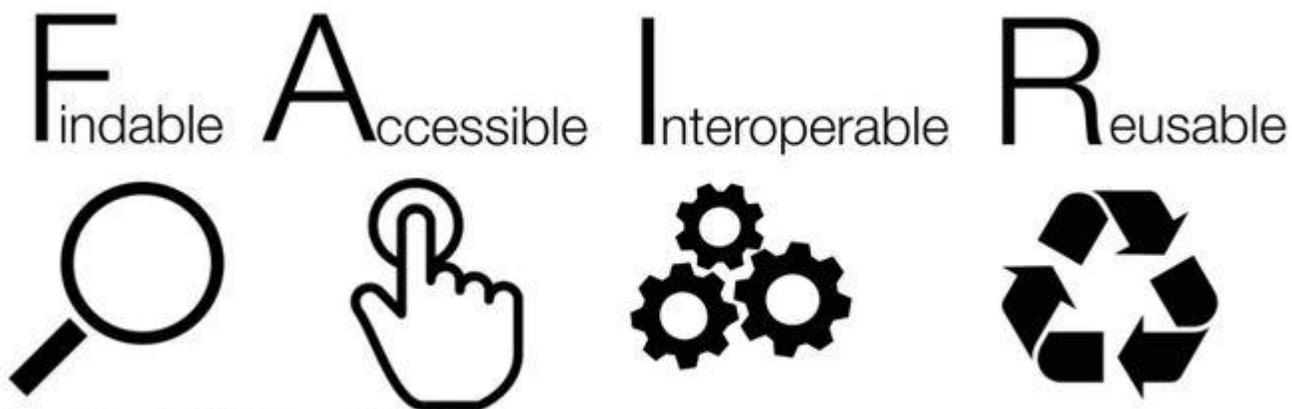


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

FAIR data – what it is & what it is not

FAIR

- refers to a set of principles with the aim of providing guidance for increasing the reusability of research data

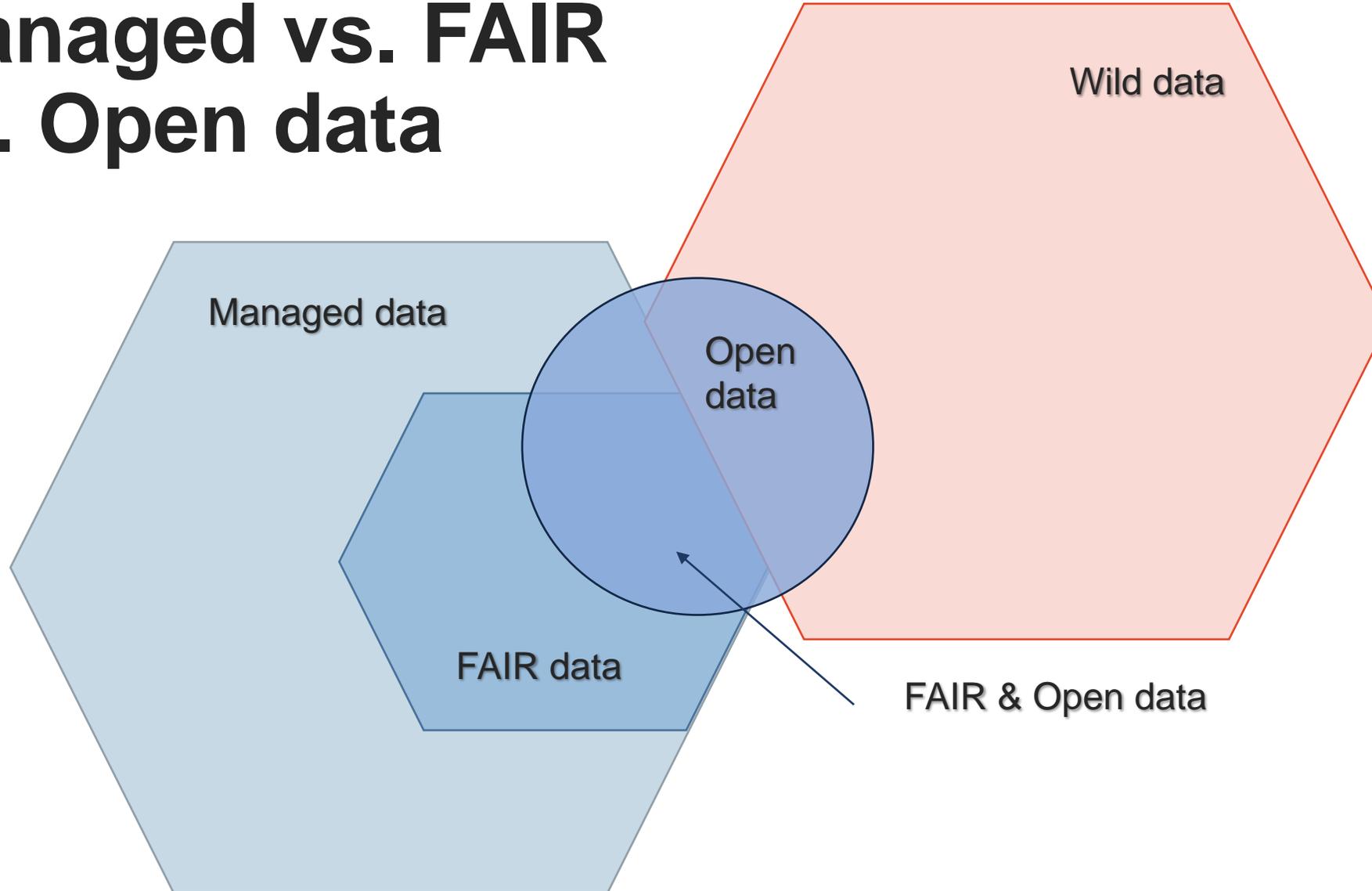
FAIR

- is not a standard
- does not require any specific technique
- is not just for humans
- is not the same as Open Data/freely accessible data.



<https://www.go-fair.org/fair-principles/>

Managed vs. FAIR vs. Open data





Findable

- Mash-up of two concepts*

Locatable = you knew it existed; you could find it.

Discoverable = you didn't know it existed but could discover it by searching.

- Dataset has a unique and **persistent identifier**.
- Data are described with rich **metadata**.
- Metadata are **registered or indexed** in a searchable resource.
 - Registered e.g., DOI metadata registered with DataCite
 - Indexed e.g., schema.org metadata indexed by Google Data

* that have various names.



Dataset has a unique and **persistent identifier**

- e.g., Digital Object Identifier (**DOI**)
- Created when data are published.
- Used when citing data.

Estimating effects of arable land-use intensity on farmland birds using joint species modelling

Citation:

Martin Stjernman, Ola Olsson, Henrik G. Smith, Ullrika Sahlin (2019). *Estimating effects of arable land-use intensity on farmland birds using joint species modelling*. Swedish National Data Service. Version 1.0.

<https://doi.org/10.5878/58w0-m352>



F
indable



Dataset has a unique and **persistent identifier**

- e.g.. I
- Rese
- A DO

Estima
modell

Citation:

Martin St
on farml
[https://doi](https://doi.org/10.5878/58w0-m352)



Selected data: 2

<https://snd.gu.se/en/catalogue/dataset/snd1080-1>

Logged in as: David Rayner | My Pages | Log out

[Find Data](#) [Manage Data](#) [Describe & Share Data](#) [News & Events](#) [About Us](#) [Contact](#)

[← Back to search results](#)

Estimating effects of arable land-use intensity on farmland birds using joint species modelling

SND-ID: snd1080-1. Version: 1.0. DOI: <https://doi.org/10.5878/58w0-m352>



Download data

SND 1080-001-V1.0.zip (170.79 KB)

Citation

Stjernman, M. et al., 2019. Estimating effects of arable land-use intensity on farmland birds using joint species modelling. Available at: <https://doi.org/10.5878/58w0-m352>.



Accessibility level

Access to data through SND
Data are freely accessible

Use of data

Things to consider when using
data shared through SND



Versions

Version 1.0. 2019-01-24



Download metadata

Wind Run Changes: The Dominant Factor Affecting Pan Evaporation Trends in Australia

D. P. RAYNER

Queensland Department of Natural Resources, Mines, and Water, Indooroopilly, Australia

(Manuscript received 18 April 2006, in final form 12 October 2006)

ABSTRACT

The Class A pan evaporation rates at many Australian observing stations have reportedly decreased between 1970 and 2000. This trend, which has increased since 2000, has become a major focus of research. The trend has been attributed to a number of factors, including changes in solar radiation, vapor pressure deficit, and wind run.

Pan evaporation trends in observed other climate variables (termed *wind run*) have been a significant step toward

Monthly anomaly time series of observed pan evaporation, modeled pan evaporation, solar radiation, vapor pressure deficit, and wind run for all stations used in the attribution study are available online (see <http://www.nrm.qld.gov.au/silo/pan/>).



Wind R

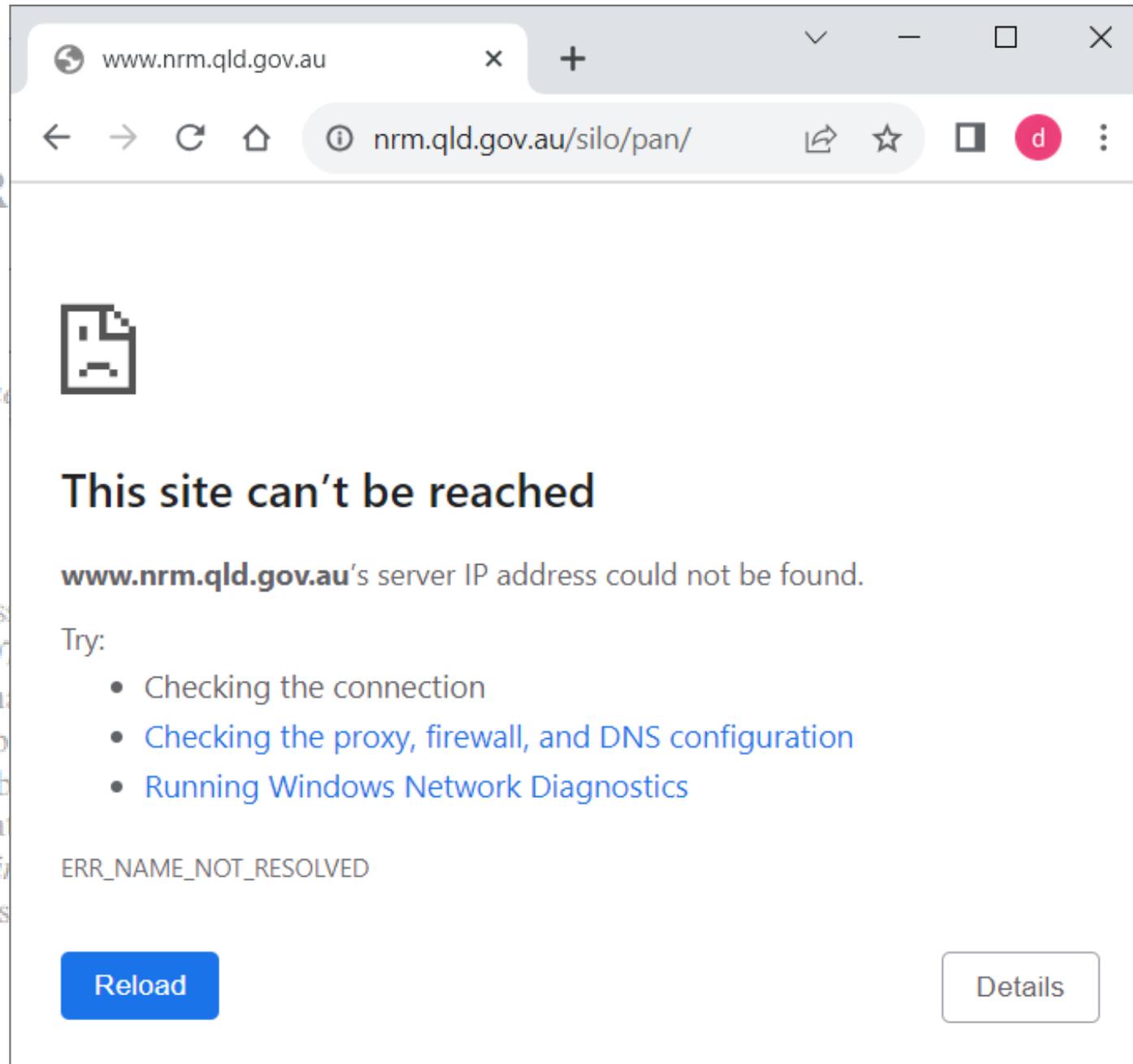
Que

The Clas
between 197
increased h
Pan evapo
trends in ob
other clima
(termed *wi*
significant s

Evaporation

, Australia

portedly decreased
have
pan evapo
tion, vapor
used in the
http://www.



The screenshot shows a web browser window with the address bar displaying `www.nrm.qld.gov.au` and the current page URL `nrm.qld.gov.au/silo/pan/`. The browser interface includes navigation buttons (back, forward, refresh, home), a search icon, and a star icon for bookmarks. The main content area displays a large error message: "This site can't be reached" with a sad face icon. Below the message, it states: "www.nrm.qld.gov.au's server IP address could not be found." Underneath, there is a "Try:" section with three bullet points: "Checking the connection", "Checking the proxy, firewall, and DNS configuration", and "Running Windows Network Diagnostics". At the bottom of the error message, the text "ERR_NAME_NOT_RESOLVED" is visible. Two buttons are present: a blue "Reload" button and a white "Details" button. A hand cursor is positioned over the "Details" button.

A_{ccessible}



Accessible

- Metadata are **retrievable from the identifier**.
- Systems built with **open standards**.
- Provide the exact conditions under which the data are accessible.

FAIR ≠ ‘**Openly accessible**’, or
‘**Freely accessible**’!

- Metadata can be accessed even if data are withdrawn.

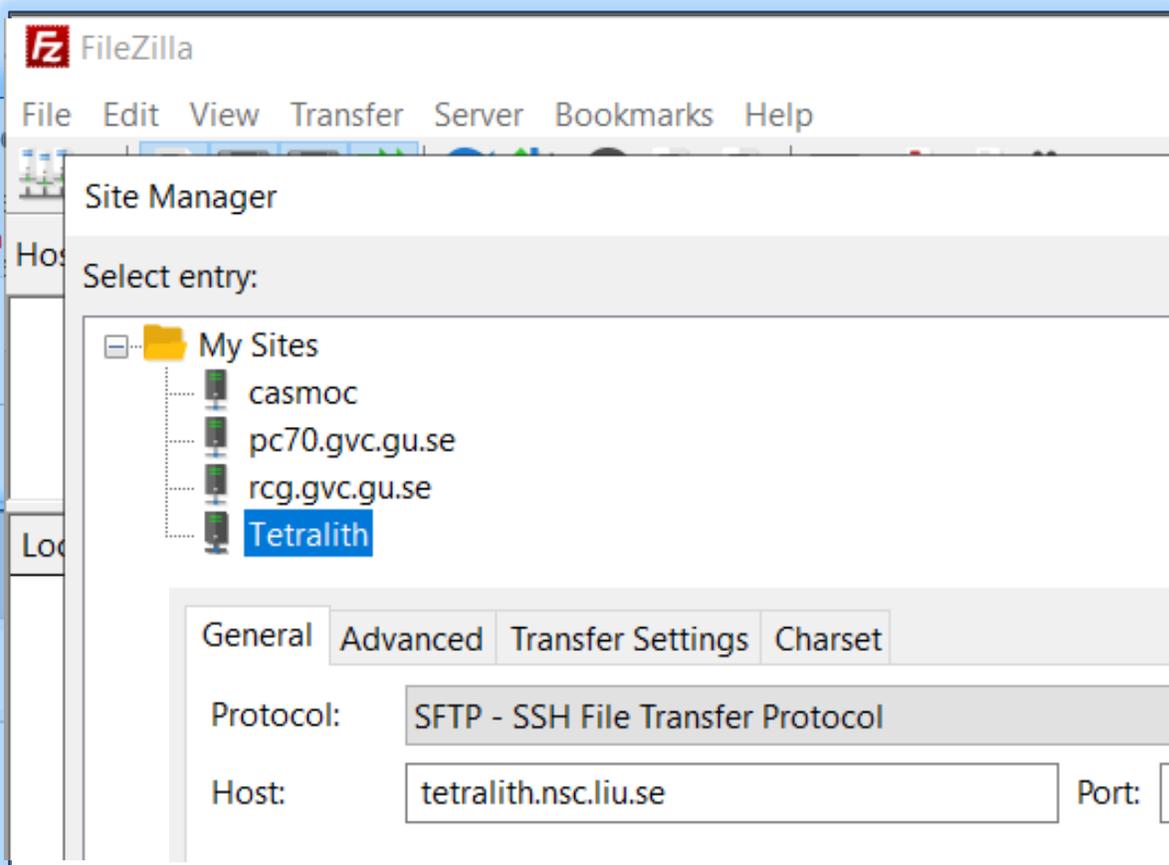
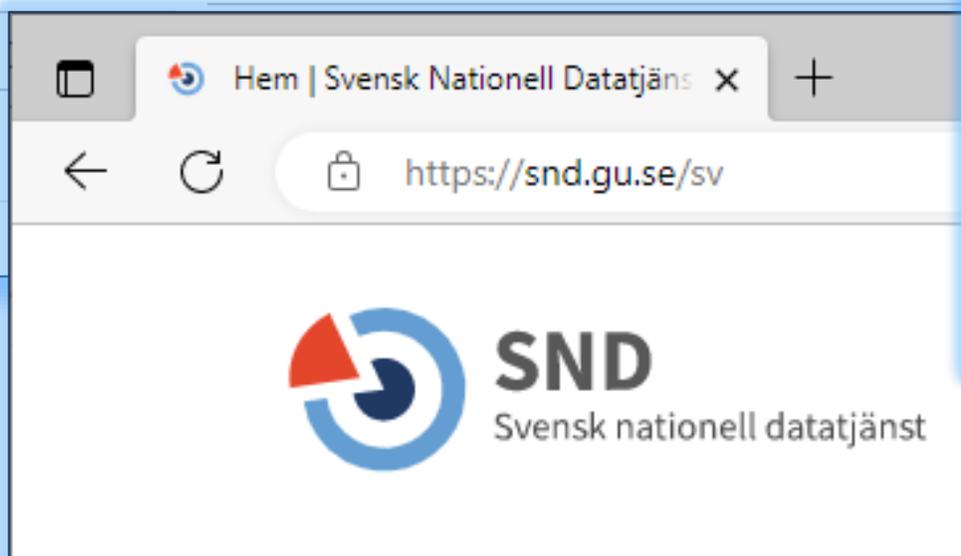
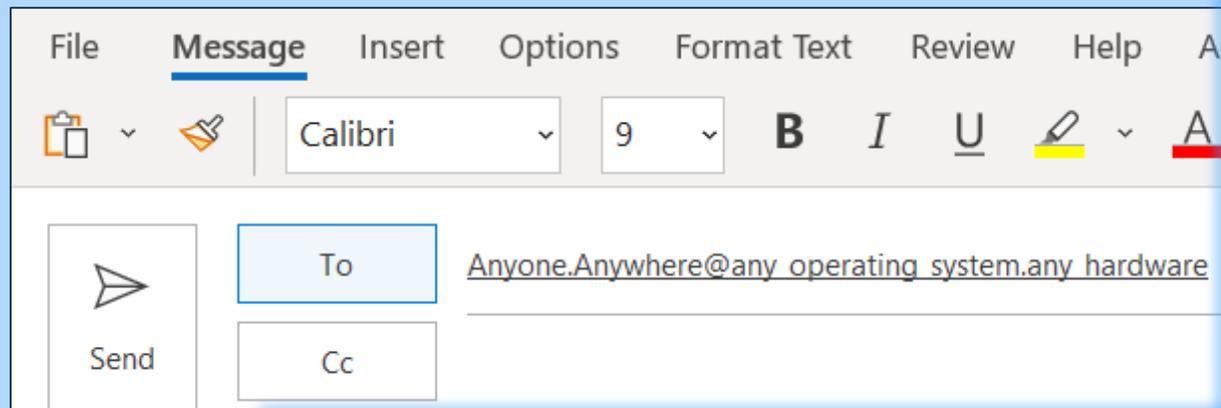


Photo by [LARAM](#) on [Unsplash](#)

Accessible



Built with open standards



(Not open standards – iMessage, Skype...)

Why You Can't Normally Use iMessage on Android

You usually can't use special end-to-end e



ies a
s the

messages
the device
iMessage
the message

Android and iOS are great in terms of features and usability. However, it is quite a difficult task for novice users to send files between Android and iOS.

Apple keeps iMessage and all its cool effects and features, including

[iMessage apps](#), exclusive to devices running iOS



By [Dvir Ben Aroya](#), January 11, 2022, 4 min read

way to drive people to buy its products. That's why there's no iMessage for Android app available on the Google Play store.

A_{ccessible}



Accessible

- Metadata are **retrievable from the identifier**.
- Systems built with **open standards**.
- Provide the exact conditions under which the data are accessible.

FAIR ≠ ‘**Openly accessible**’, or
‘**Freely accessible**’!

- Metadata can be accessed even if data are withdrawn.

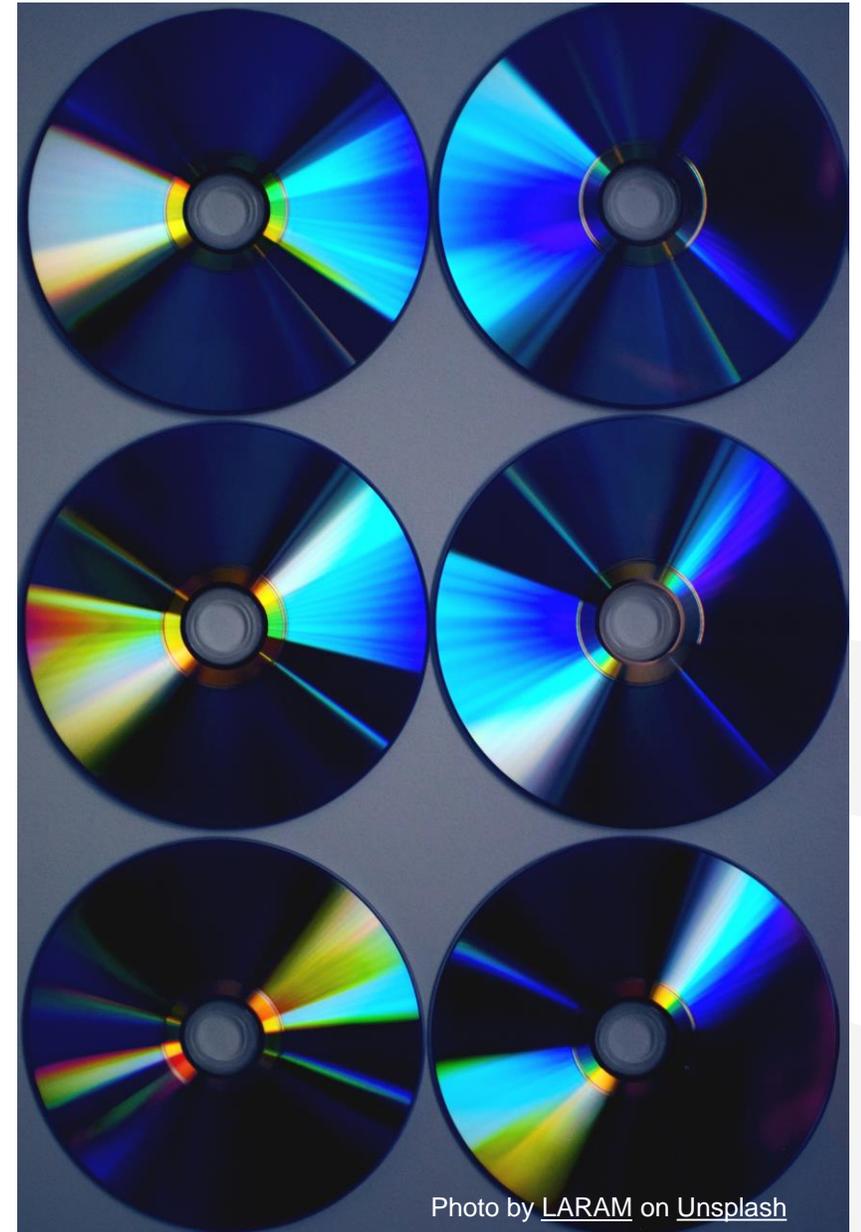
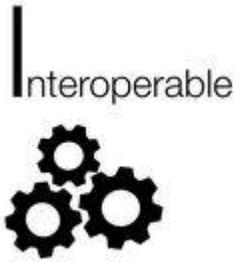


Photo by [LARAM](#) on [Unsplash](#)



Interoperable

- Provide a “**common understanding**” of digital objects.
- **Link** to other information.



Interoperable

Off-peak frequency in regional public transport

SND-ID: 2021-314

Overview

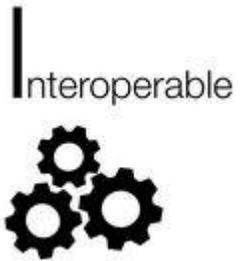
Data and documentation

Creator/Principal investigator(s)

[Joel Hansson](#) - Lund University, Department of Technology and Society 

Description

The purpose of this study is to look into patronage effects of extended supply outside peak hours on regional public transport services. The patronage effects are explored through four case studies from the region of Scania (Skåne), south Sweden. The cases include rail, coach, and bus services where substantial improvements have been made outside peak hours, resulting in at least hourly all-day services. Two datasets are used: one with panel data, covering 10-14 years of annual measurements of the daily number of trips, and one with before-and-after data where the number of trips on individual departures have been recorded at two points in time.



Interoperable

The screenshot shows the DataCite Commons interface. At the top left is the DataCite Commons logo. A search bar contains the text 'Type to search...'. To the right are links for 'Pages', 'Support', and a 'Sign In' button. Below the search bar are navigation tabs for 'Works', 'People', 'Organizations', and 'Repositories', with 'Works' being the active tab. The main content area features the title 'Off-peak frequency in regional public transport' in blue, followed by a Creative Commons license icon (CC BY) and the DOI 'https://doi.org/10.5878/xr4y-0143'. On the left side, there are two buttons: 'Add to ORCID Record' and 'Download Metadata'. Below these is a 'Cite as' section with the citation: 'Hansson, J. (2021). *Off-peak frequency in regional public transport* (Version 1) [Data set]. Lund University. <https://doi.org/10.5878/XR4Y-0143>'. A dropdown menu shows 'APA' as the selected citation style. Below the citation is a 'Share' section. The description of the work is provided in the main content area, starting with 'The purpose of this study is to look into patronage effects of extended supply outside peak hours on regional public transport services...'. At the bottom of the page, a footer note reads 'Version 1 of Content published 2021 in Swedish National Data Service'.



Interoperable

Interoperable

Off-peak frequency in regional public transport

SND-ID: 2021-314

Overview

Data and documentation

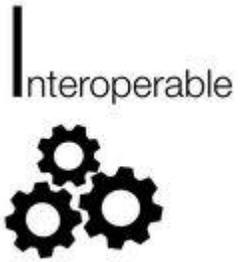
Creator/Principal investigator(s)

[Joel Hansson](#) - Lund University, Department of Technology and Society 

orcid.org/0000-0001-6478-0091

Description

The purpose of this study is to look into patronage effects of extended  outside peak hours on regional public transport services. The patronage effects are explored through four case studies from the region of Scania (Skåne), south Sweden. The cases include rail, coach, and bus services where substantial improvements have been made outside peak hours, resulting in at least hourly all-day services. Two datasets are used: one with panel data, covering 10-14 years of annual measurements of the daily number of trips, and one with before-and-after data where the number of trips on individual departures have been recorded at two points in time.



Metadata can be **linked to other information**

ORCID
Connecting research and researchers

SIGN IN/REGISTER English

Search...

iD
https://orcid.org/
0000-0001-6478-0091

Is this you? [Sign in to start editing](#) Printable version

Name
Joel Hansson

Works (7) Sort

Patronage effects of off-peak service improvements in regional public transport

European Transport Research Review
2022-12 | Journal article [Show more detail](#)
DOI: [10.1186/s12544-022-00543-4](https://doi.org/10.1186/s12544-022-00543-4)
CONTRIBUTORS: Joel Hansson; Fredrik Pettersson-Löfstedt; Helena Svensson; Anders Wretstrand

Source: Crossref

Effects of rural bus stops on travel time and reliability



R_{eusable}



Reusable

- “**Richly described**” - **documentation** so user can decide if the data are useful in a particular context.
- Clear **usage rights** – what is a user allowed do with it?

Reusable



Clear usage rights – license

Off-peak frequency in regional public transport

SND-ID: 2021-314

Overview

Data and documentation

Data
off

License

[Creative Commons Public Domain Dedication \(CC0 1.0\)](#)

Offpeak regional PT data.xlsx (70.97 KB)

Annual trip data.csv (6.62 KB)

Variables.csv (1.77 KB)

Offpeak variables.xlsx

Offpeak variables.csv





Usage rights – license



CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

This is a human-readable summary of the [Legal Code \(read the full text\)](#). [Disclaimer](#)

No Copyright



The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.



FAIR Principles – Summary

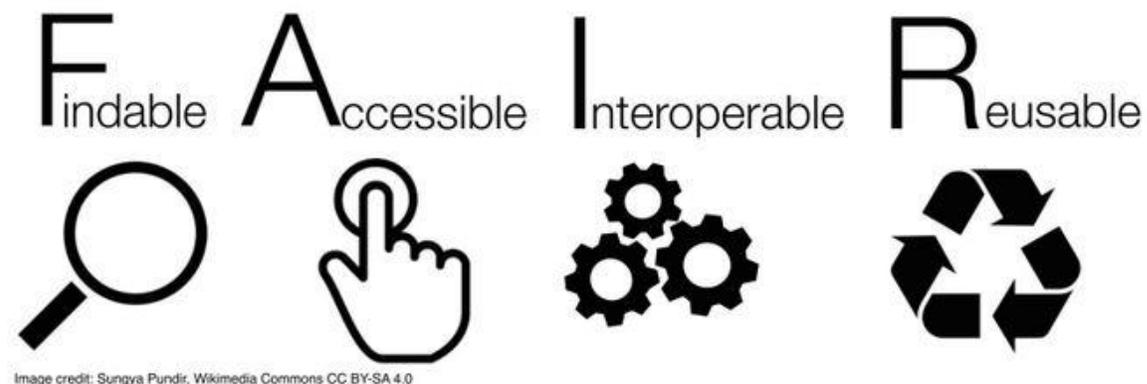
- The “*gold standard*” for sharing data.

- Most important points in FAIR:

- **persistent identifier** e.g., DOI
- structured **metadata**
- **documentation**
- clear usage **license**

- ***Important!*** FAIR data can have limited access!

... as long as there are clear instructions on how to access the data!



How do I create FAIR research data?

- **Use a repository** (“public archive”)!
 - These do **not*** count as FAIR:
 - University’s cloud storage
 - Supplement to articles.

GUbox will be discontinued

NEWS: JUN 17, 2021

The contract for GUbox expires in the end of 2021 and the service will therefore be discontinued sometime in the first months of 2022. All files currently stored in GUbox must be moved.

The schedule previously published on the toolpage for GUbox here in the Staff portal, where 31st of August was the last date to move material yourself, does not apply anymore.

* usually

Repositories (“public archives”)



SND

Swedish National Data Service

We accept data from all disciplines.



cessda
DC Data Catalogue

[Consortium of European Social Science Data Archives](#)



A free and open platform for validating and sharing
BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data



Data management plans

What is **data management**?

- **"Data management"**
- *A systematic and planned organization and structuring of how research material is handled during the research process.*



Image: [Nick Youngson CC BY-SA 3.0](#)

Data management is a journey

Work with data



Finish and make accessible



Plan



What is a Data Management Plan (DMP)?

- Tool that facilitates systematic data handling.
- A framework for how data should be handled.
- Starts as a planning tool...
- ...becomes a journal of actions and decisions...
- ... when finalized, a reference guide for your data.



What should you update in your **DMP**?

Plan

Plan

- Ethical approvals
- Personal data
- Information classification
- Agreements

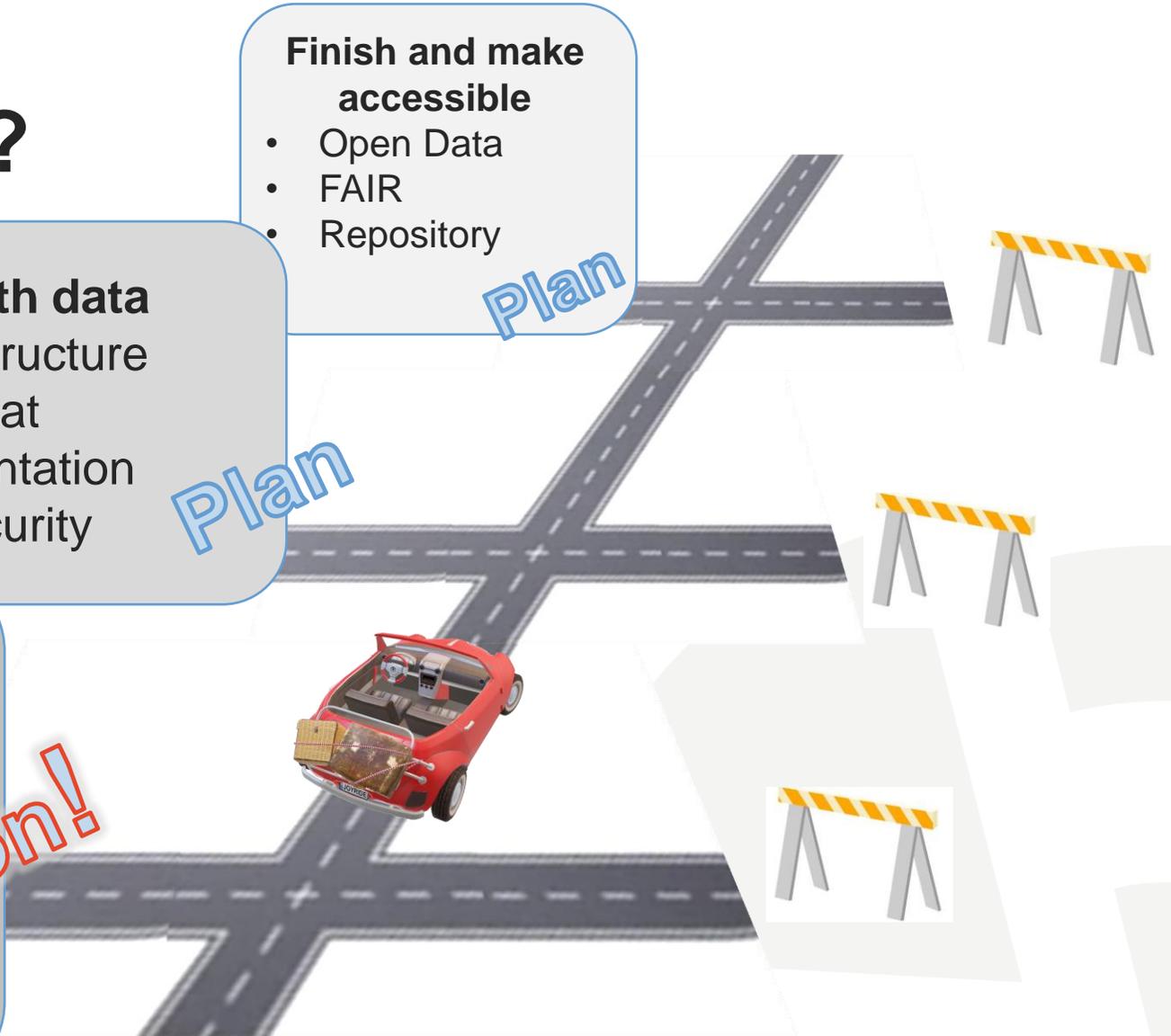
Documentation!

Work with data

- Folder structure
- File format
- Documentation
- Data security

Finish and make accessible

- Open Data
- FAIR
- Repository



Why create a DMP?

- Simplify data management
- Think early
- Ensure that the material is well handled.
- Facilitates the preparation of the material for archiving and sharing.
- Some funders requires a DMP!! (H2020, Research Council, Formas...)

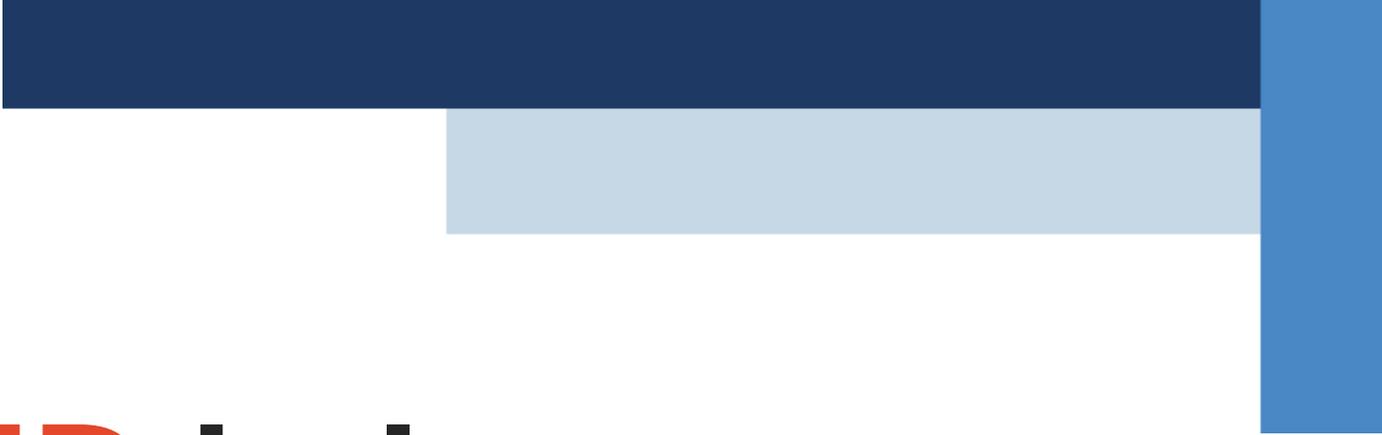


SND's Checklist for Data Management Plan

- Help you design a data management plan.
- Identify which parts of data management are important to your research project.
- Can be used for a complete plan for an entire project.

*All the advice in this presentation
is included in SND's Checklist*

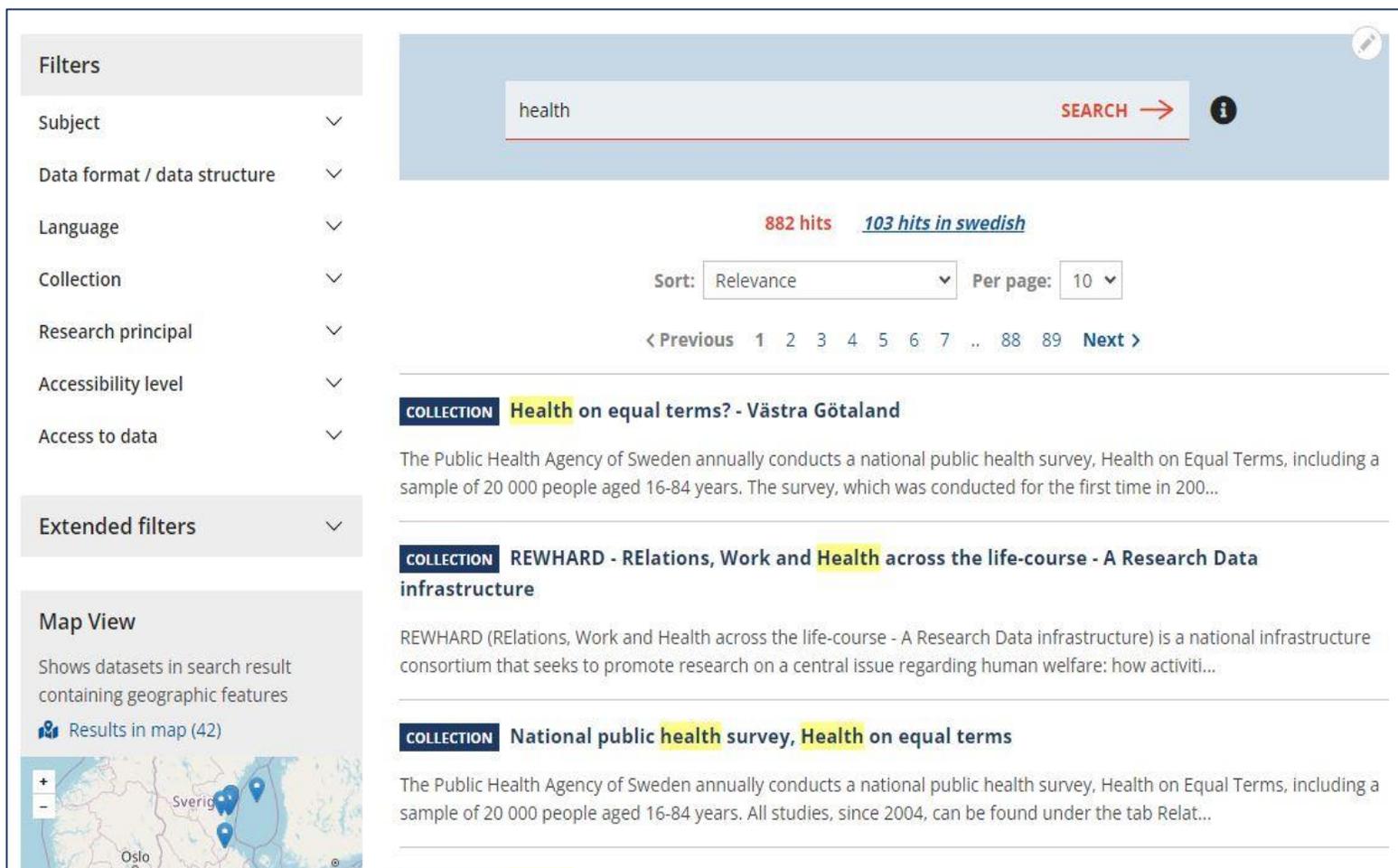




How does **SND help
researchers make research
data available for reuse?**

Research Data Catalogue

Search and find data



The screenshot displays the search interface of the Research Data Catalogue. On the left, there is a sidebar with filter categories: Filters, Extended filters, and Map View. The main search area includes a search bar with the term 'health', a search button, and an information icon. Below the search bar, it shows 882 hits, with 103 hits in Swedish. The results are sorted by Relevance and displayed 10 per page. The pagination shows results 1 through 89, with 'Previous' and 'Next' buttons. Three search results are visible, each starting with a 'COLLECTION' label and a title. The first result is 'Health on equal terms? - Västra Götaland', the second is 'REWHARD - Relations, Work and Health across the life-course - A Research Data infrastructure', and the third is 'National public health survey, Health on equal terms'. Each result includes a brief description of the dataset.

Filters

- Subject
- Data format / data structure
- Language
- Collection
- Research principal
- Accessibility level
- Access to data

Extended filters

Map View

Shows datasets in search result containing geographic features

 Results in map (42)

Search

health **SEARCH** 

882 hits [103 hits in swedish](#)

Sort: Relevance Per page: 10

[< Previous](#) 1 2 3 4 5 6 7 .. 88 89 [Next >](#)

COLLECTION **Health on equal terms? - Västra Götaland**

The Public Health Agency of Sweden annually conducts a national public health survey, Health on Equal Terms, including a sample of 20 000 people aged 16-84 years. The survey, which was conducted for the first time in 200...

COLLECTION **REWHARD - Relations, Work and Health across the life-course - A Research Data infrastructure**

REWHARD (Relations, Work and Health across the life-course - A Research Data infrastructure) is a national infrastructure consortium that seeks to promote research on a central issue regarding human welfare: how activiti...

COLLECTION **National public health survey, Health on equal terms**

The Public Health Agency of Sweden annually conducts a national public health survey, Health on Equal Terms, including a sample of 20 000 people aged 16-84 years. All studies, since 2004, can be found under the tab Relat...

Research Data Catalogue

 Search and find data



Share data via
the SND
catalogue

Data from Opinionskoll 2017. - Data from third Opinionskoll survey.

2022-159-1

 Editing

Research area profile

Social Sciences

 Change

Sections

- 1. Files and accessibility >
- 2. Citation and description >
- 3. Administrative information >
- 4. Collection and method >
- 5. Topics and keywords >
- 6. Geographic coverage >
- 7. Publications and relations >

1. Files and accessibility

Files

Click 'Select data files' to upload the data files you want to make available via SND. You can upload multiple files at a maximum upload size of 500 MB per file. If you plan on uploading larger files or sensitive data, please contact incoming@snd.gu.se and we will assist you.

If you are creating a new version of an already published dataset, you need to upload all the data files that should be included in this version. [View instructions for creating a new version of a dataset.](#)

Show more ▾

File name

 Övningsdata BAS Online_Data.xlsx

38.1 kB



Select data files 

Documentation files

When research data are made available, it is also important to share the documentation necessary for other researchers to understand and use the data. Examples include variable lists, readme files, and project documentation. It is particularly

Research Data Catalogue

 Search and find data

 Share data via the SND catalogue



Share Data: Step by Step

Data files

- At present, data shared through the SND research data catalogue can be accessed via the following links: [Data files](#) (with the following exception: [Data files](#))

Metadata

Metadata is structured information used to describe and categorize data.

- Data files are shared through the SND research data catalogue

Documentation

Relevant documentation **must** be applied to ensure that users can understand and reuse the data. Give context and understanding of the data.

It can be:

- When data files are shared, relevant documentation must be applied to ensure that users can understand and reuse the data. Give context and understanding of the data.
- The name of the data files must be descriptive and understandable as a context for the data.
- Variable lists with explanations of the variables
- Questionnaires or surveys
- Interview forms, including interview guides
- Code lists and code books

Manage data

- Data management information, advice, and best practices.
- Data management plan
- Guides to good research data practices for different types of data

Manage data

Data management describes the handling, organisation, and structuring of research material during the research process. On these pages we have collected useful information about how to maintain, preserve and make research data accessible.



Plan

There is good reason to start planning for how you will collect and manage data in the research project as early as possible. Here you can find information about storage, security, costs, and other factors that you want to consider at the start of a project.



Organise

Data should be organised in a way that makes it easy to find the right version of the right file at the right time. Here you can read about what to consider when you create a folder structure, give names to and manage versions of your files, and choose file formats.



Data management plan

It is getting increasingly common with requirements for data management plans for research projects. Here we describe what a data management plan, or DMP, is and what information it may contain. You can also read and download the SND DMP Checklist.

[Read more about data management plans >](#)

→ [Guides to good research data](#)

Good Luck!

David Rayner
Lisa Isaksson

david.rayner@snd.gu.se
lisa.isaksson@snd.gu.se

SND.se



SND

Swedish National Data Service

| University of Gothenburg - Chalmers University of Technology - Karolinska Institutet - KTH Royal Institute of Technology - Lund University - Stockholm University - Swedish University of Agricultural Sciences - Umeå University - Uppsala University