This work is licensed under a Creative Commons Attribution 4.0 International License

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.

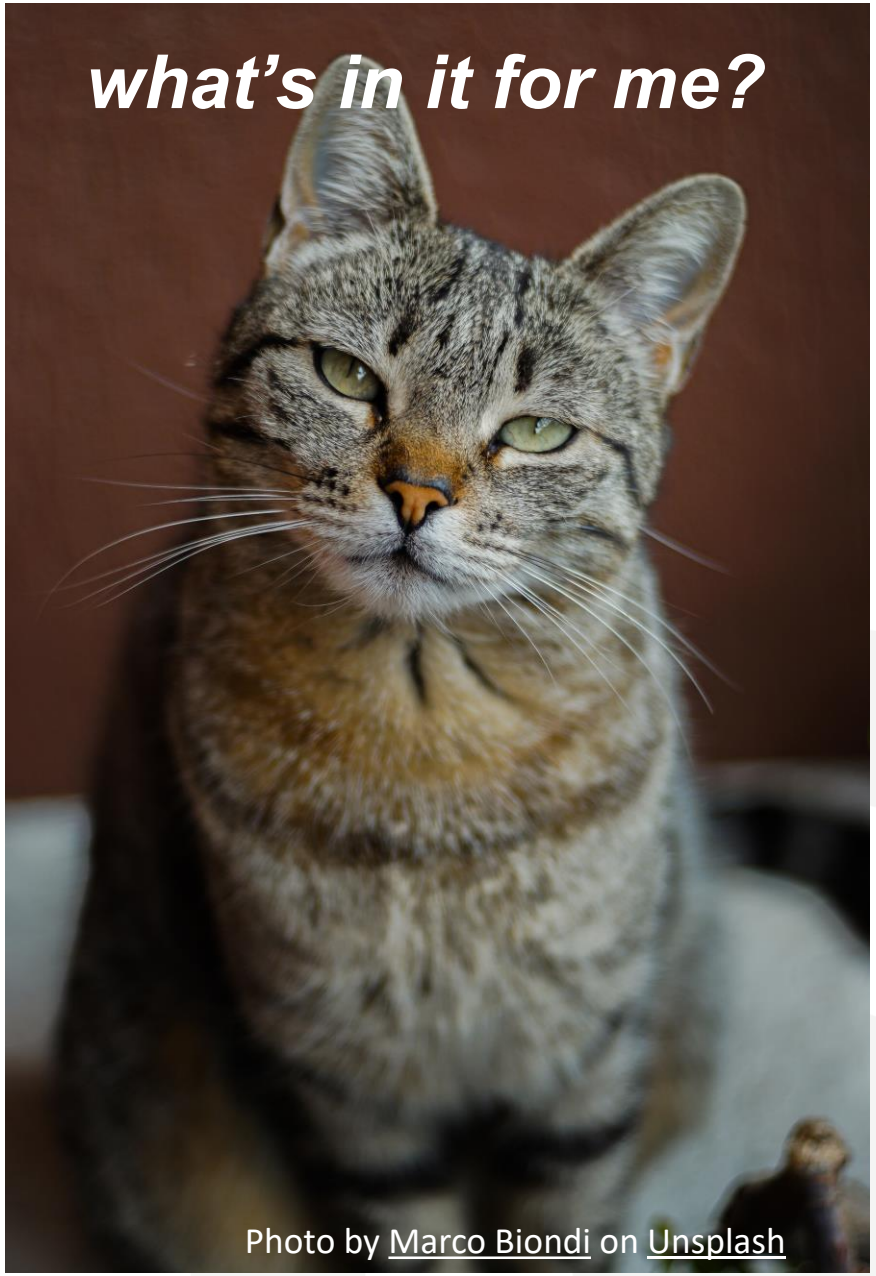


Photo by Marco Biondi on Unsplash

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

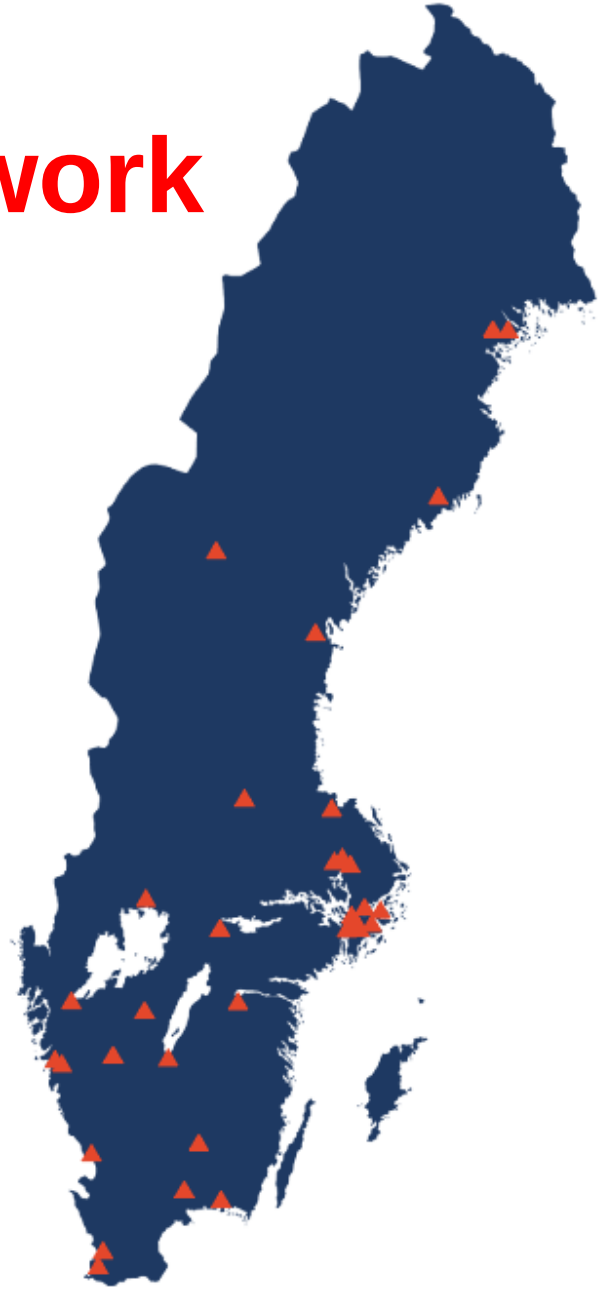
Swedish National Data Service, SND

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**.**

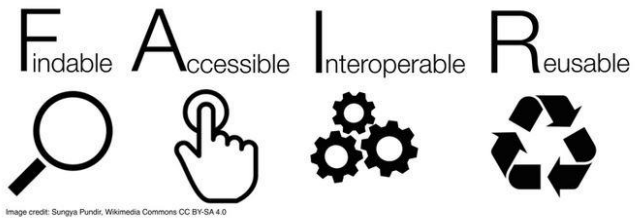


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

“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

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

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.

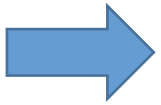
- Suppression/masking – remove!
- Tokenization
  - Replacing identifiers consistently.
- Generalization – reduce granularity
  - Categorizing age into age groups.
  - Recoding rare categories as “other”
  - Replacing address with the name of town/municipality/region

Anonymizing (de-identifying) data

- Interpretation differs within EU

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

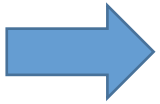
! Archiving or replication regulations may mean you *must* retain original dataset!



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*?



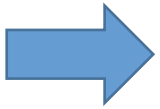
Freely accessible

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**.**

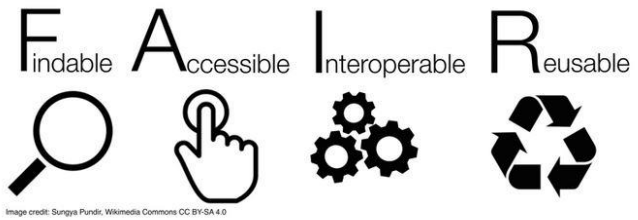


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

FAIR principles

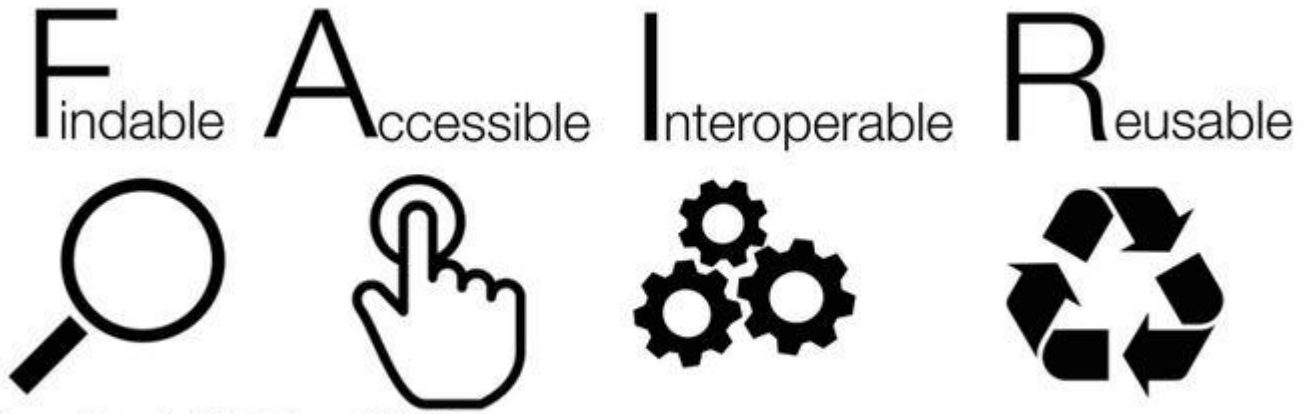


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

FAIR data – what it is & what it is not

FAIR

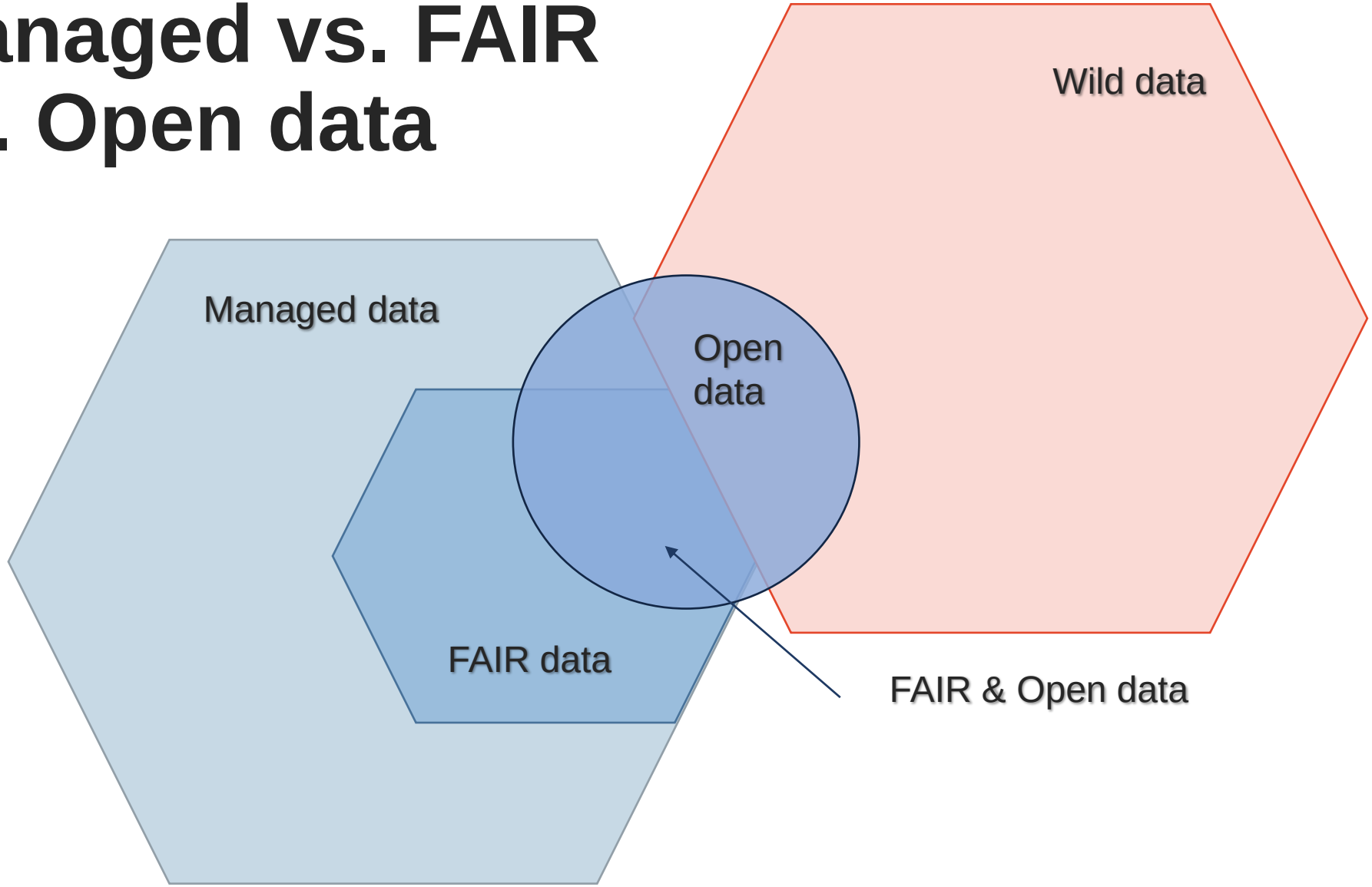
- refers to a set of principles with the aim of providing guidance for increasing the reusability of research data

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

Findable

Dataset has a unique and persistent identifier

- e.g.. D
- Resea
- A DO

SND
Swedish National Data Service

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

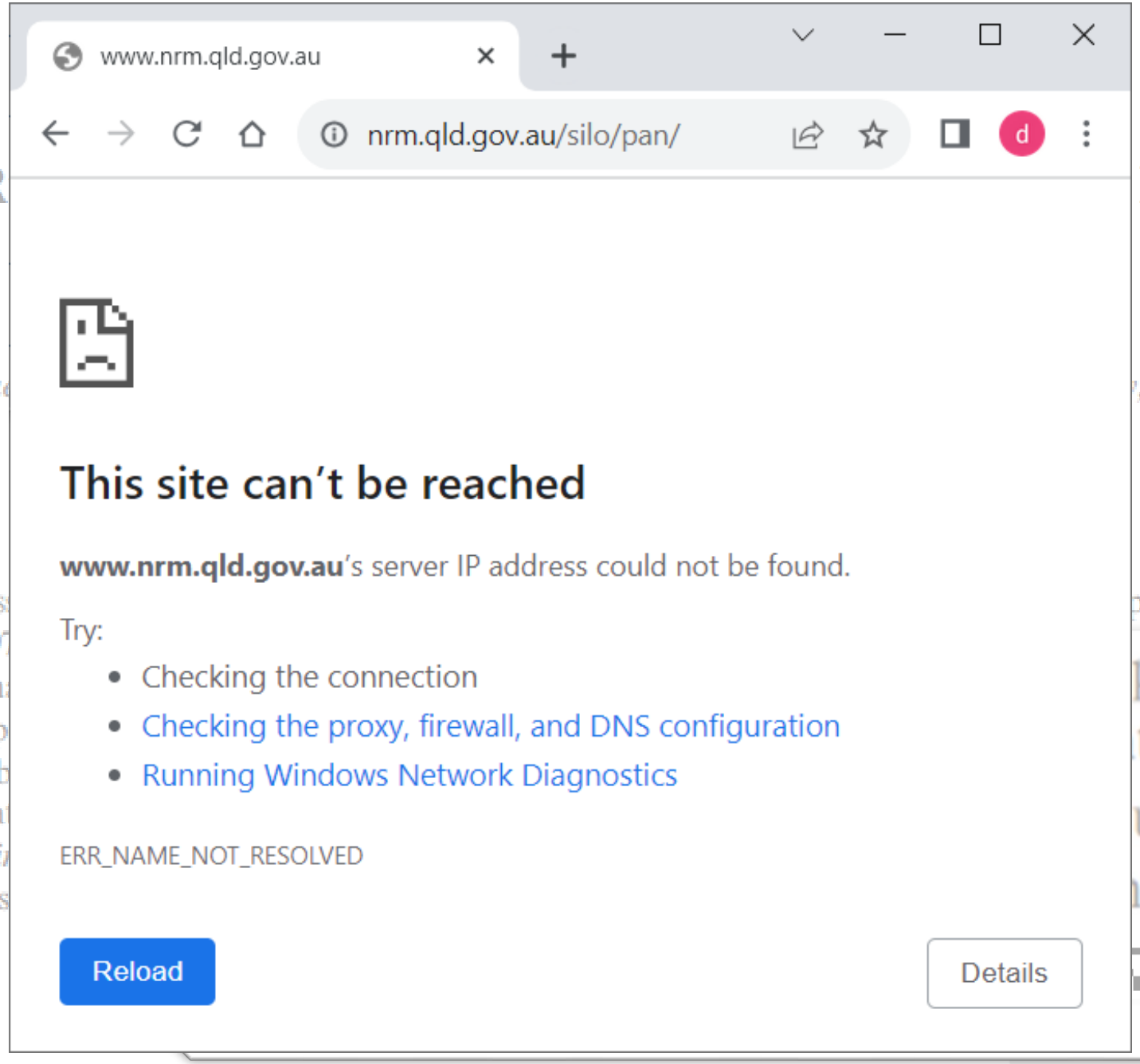
(Manuscript received 18 April 2006, in final form 12 October 2006)

ABSTRACT

The Class A pan evaporation rates at many Australian observing stations have reportedly decreased between 1970 and 2 have increased has beco

Pan evaporation trends in observed other climate varia (termed *wind run*) significant step tow

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/).


www.nrm.qld.gov.au
nrm.qld.gov.au/silo/pan/
This site can't be reached
www.nrm.qld.gov.au's server IP address could not be found.
Try:
Checking the connection
Checking the proxy, firewall, and DNS configuration
Running Windows Network Diagnostics
ERR_NAME_NOT_RESOLVED
Reload
Details

Accessible

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

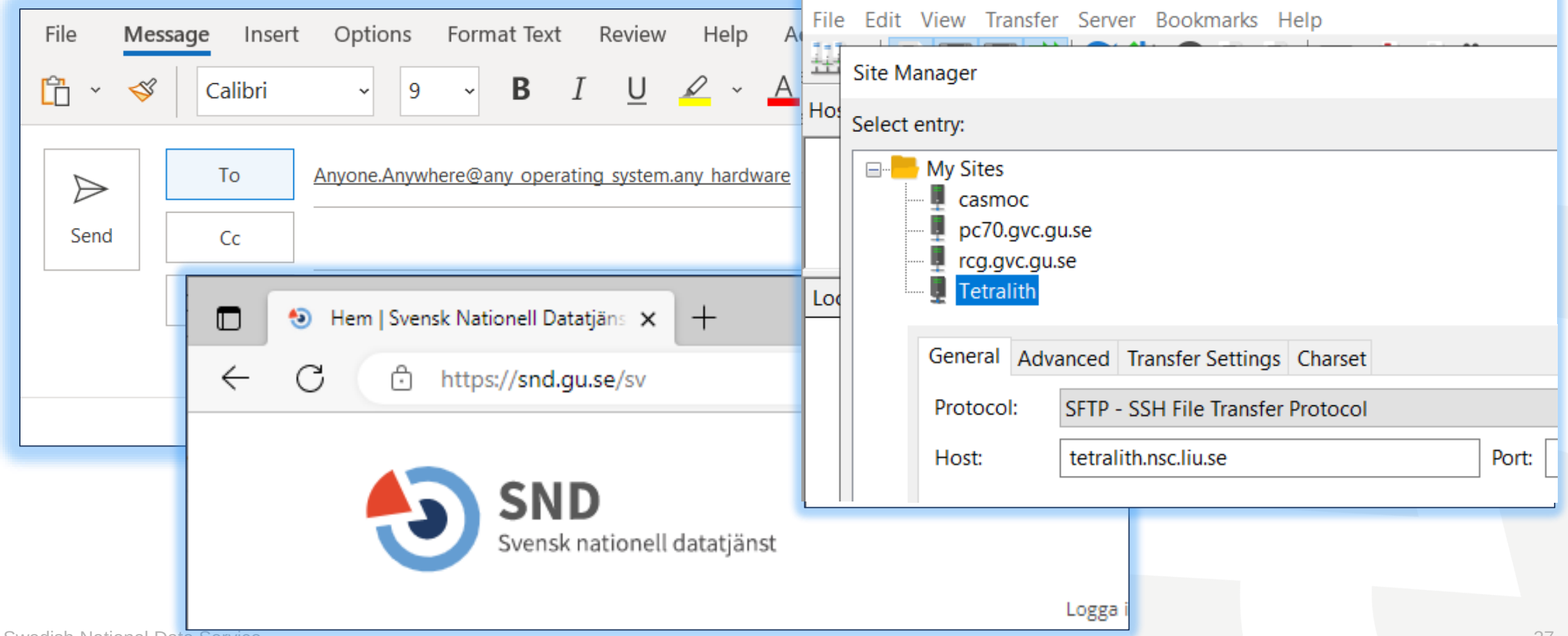
FAIR ≠ ‘Openly accessible’, or
‘Freely accessible”!

- 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 ses a special end-to-end e s the messages the device iMessage the messa

Apple keeps iMessage and all its cool effects and features, including iMessage apps, exclusive to devices running iOS way to drive people to buy its products. That's why there's no iMessage for Android app available on the Google Play store.

By **Dvir Ben Aroya**, January 11, 2022, 4 min read

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.

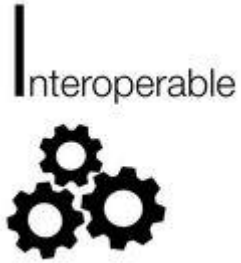
Accessible

- 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

Interoperable

Off-peak frequency in regional public transport

SND-ID: 2021-314

Overview | Data and documentation

Creator/Principal investigator(s)

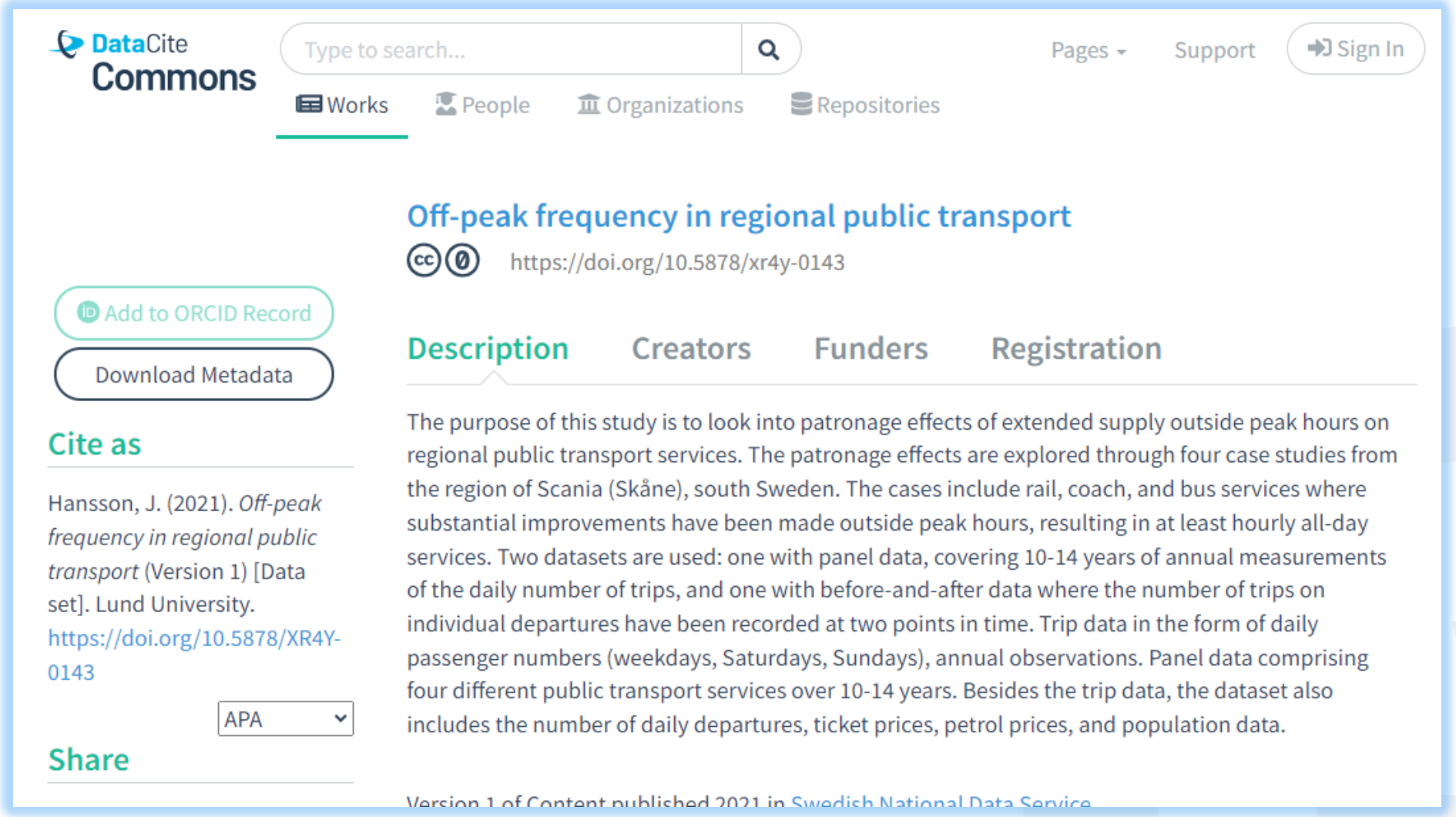
Joel Hansson - Lund University, Department of Technology and Society

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

Interoperable



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

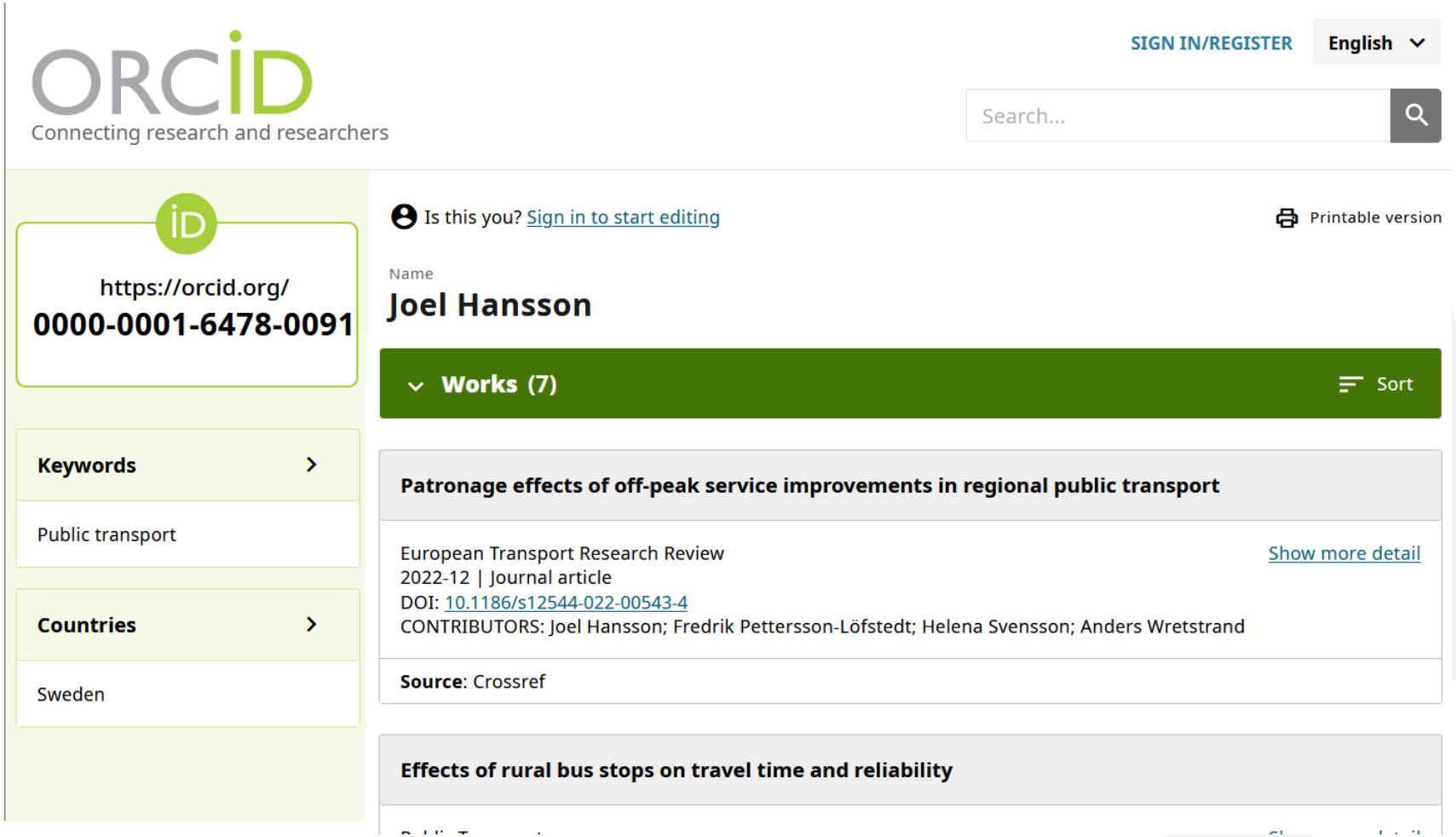
orcid.org/0000-0001-6478-0091

Description

The purpose of this study is to look into patronage effects of extended [illegible] 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

Metadata can be linked to other information

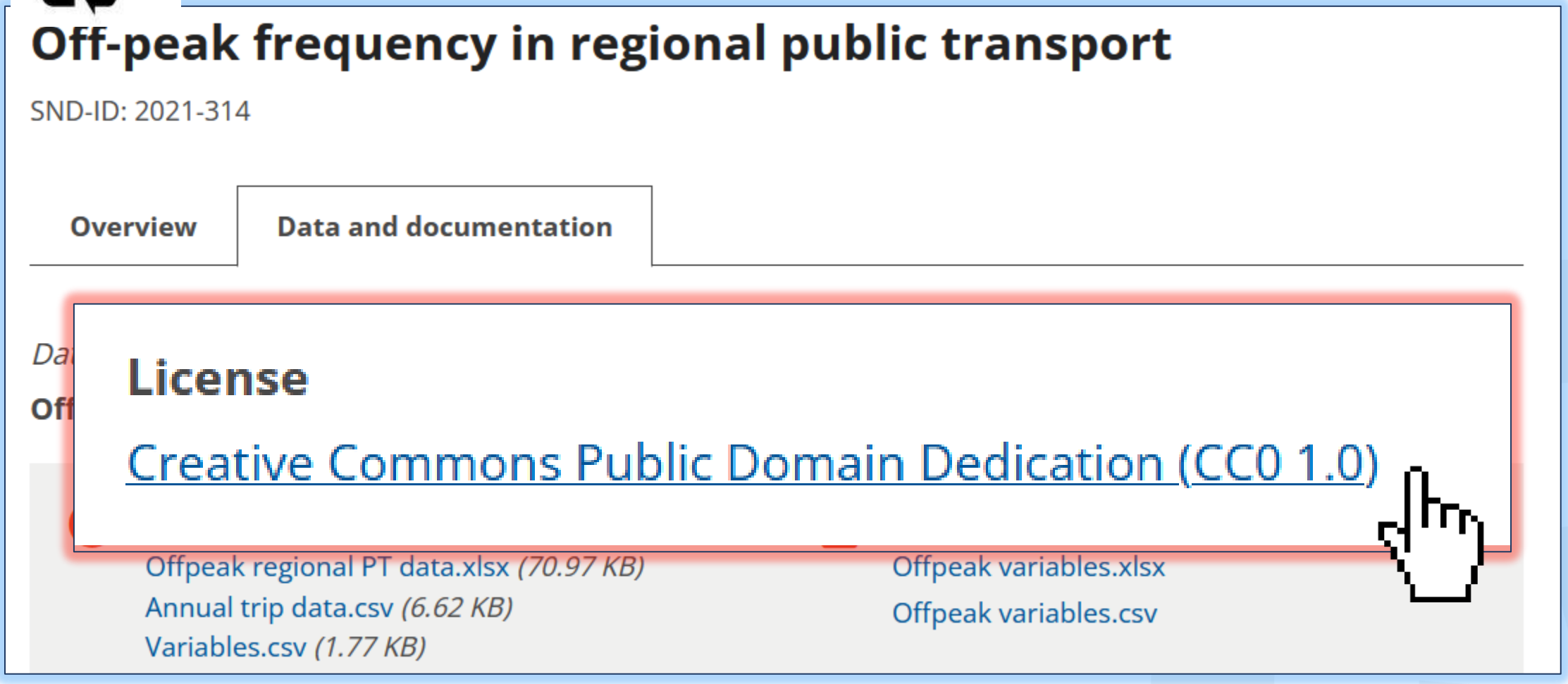


Reusable

- “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



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

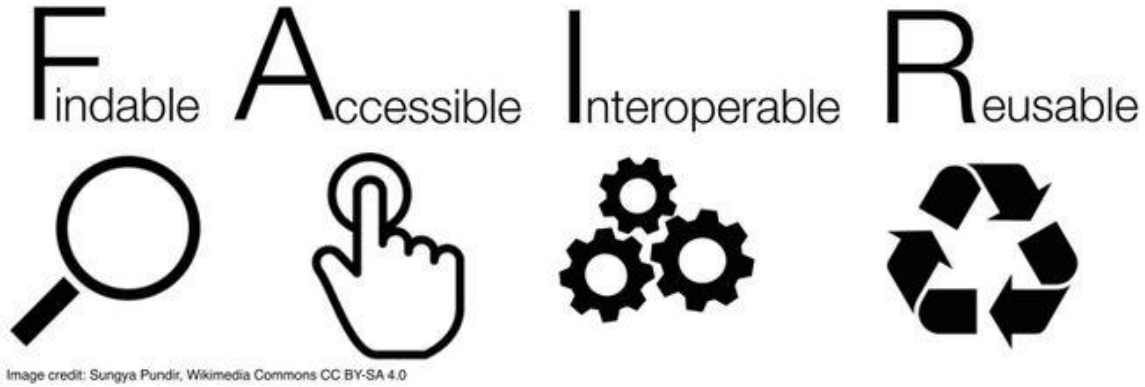


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

- *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	OpenNEURO **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?

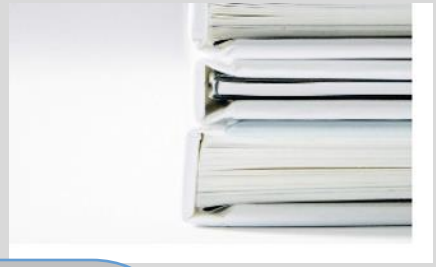
- **”Data management”**
- *A systematic and planned organization and structuring of how research material is handled during the research process.*

Image: Nick Youngson CC BY-SA 3.0

Data management is a journey

Plan

Work with data



Finish and make accessible

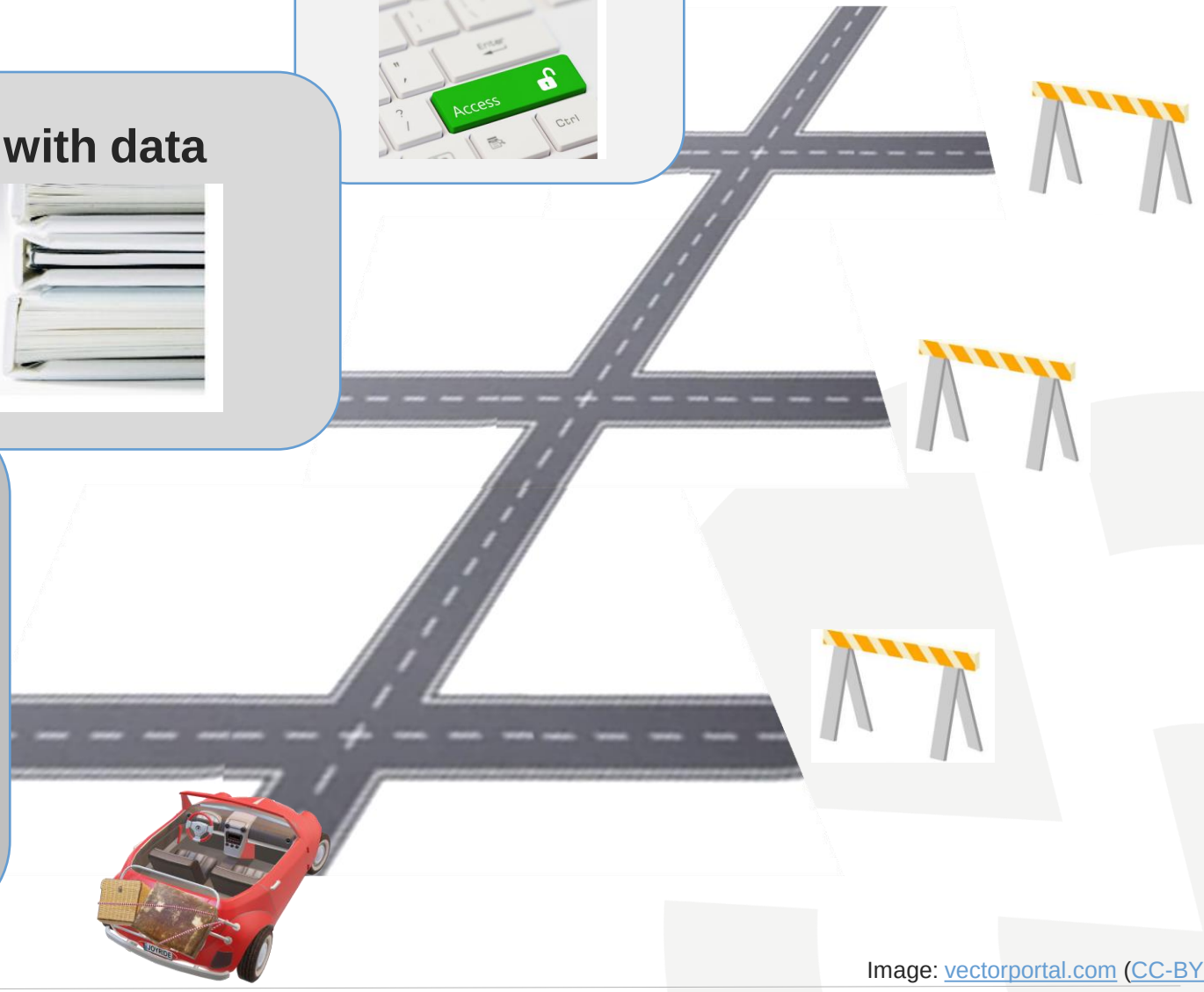


Image: vectorportal.com (CC-BY 4.0)

What is a Data Management Plan (DMP)?

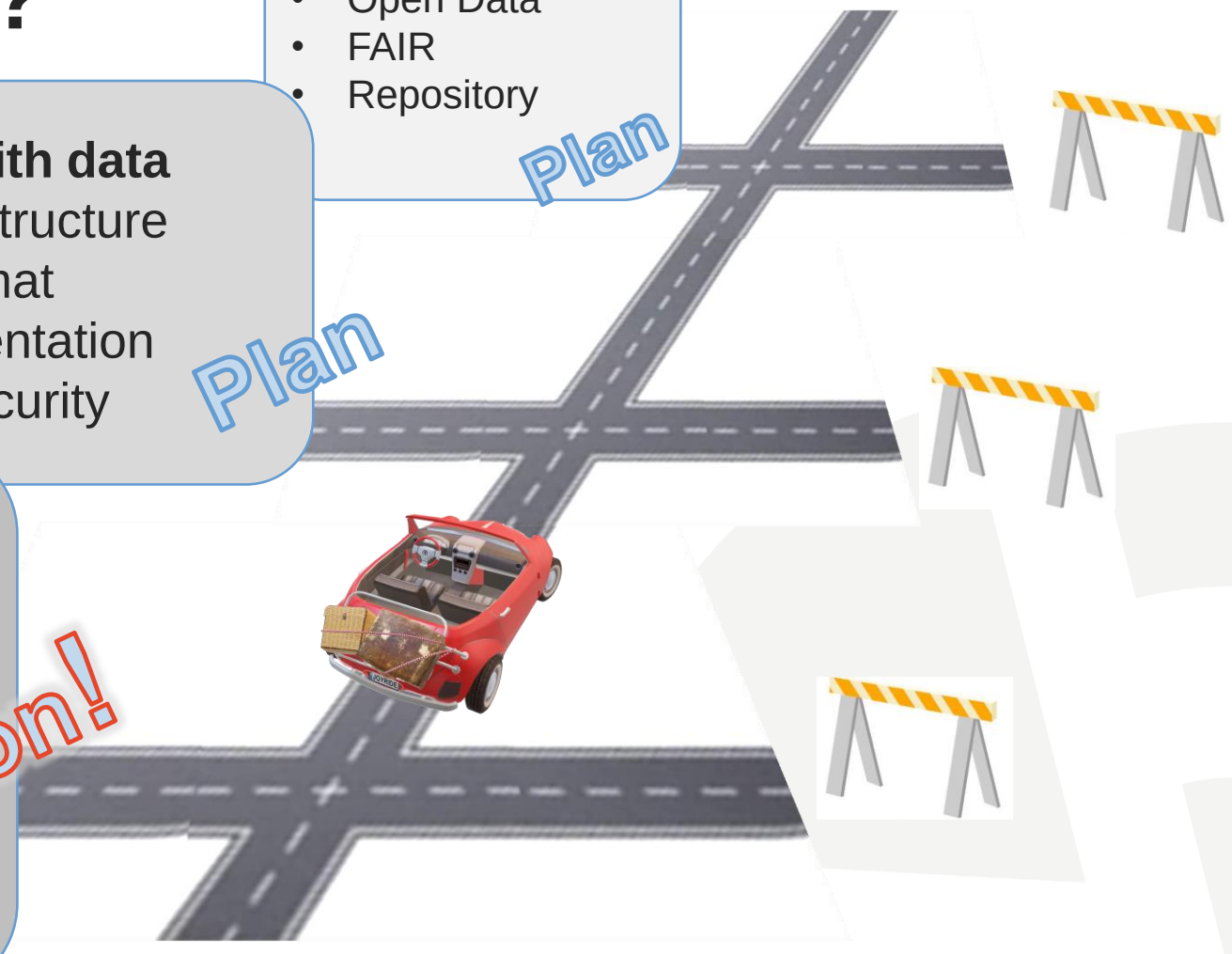
- Tool that facilitates systematic data handling.
- A framework for how data should be handled.
- Starts as a planning tool…
- …becomes a journal of actions and decisions…
- … when finalized, a reference guide for your data.

What should you update in your DMP?

Finish and make accessible
- Open Data
- FAIR
- Repository

Work with data
- Folder structure
- File format
- Documentation
- Data security

Plan
- Ethical approvals
- Personal data
- Information classification
- Agreements



Plan

Image: vectorportal.com (CC-BY 4.0)

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…)

Image: Juhan Sonin (CC By 2.0)

SND’s Checklist for Data Management Plan

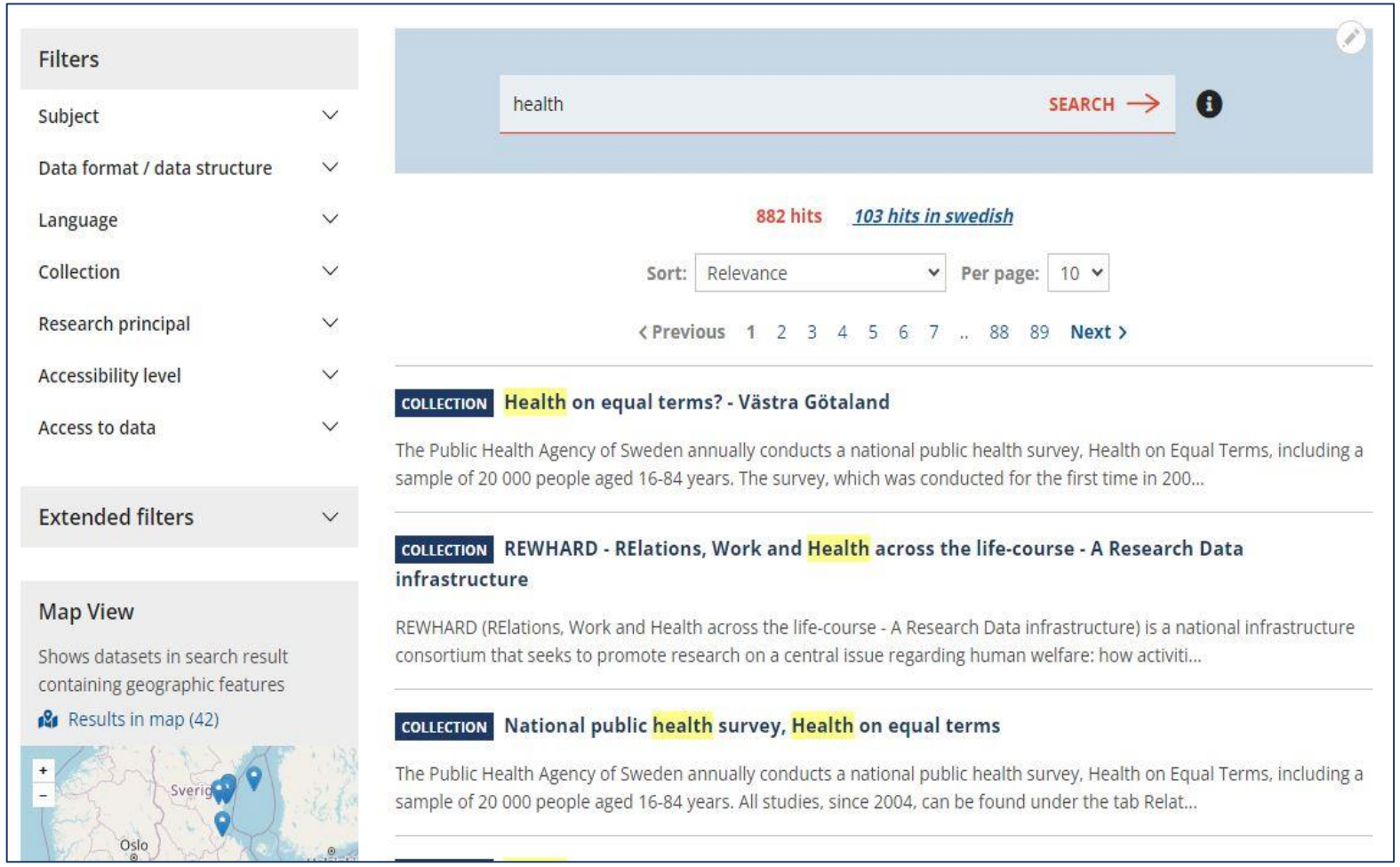
- Help you design a data management plan.
- Identify which parts of data management are important to your research project.
- Can be used for a complete plan for an entire project.

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

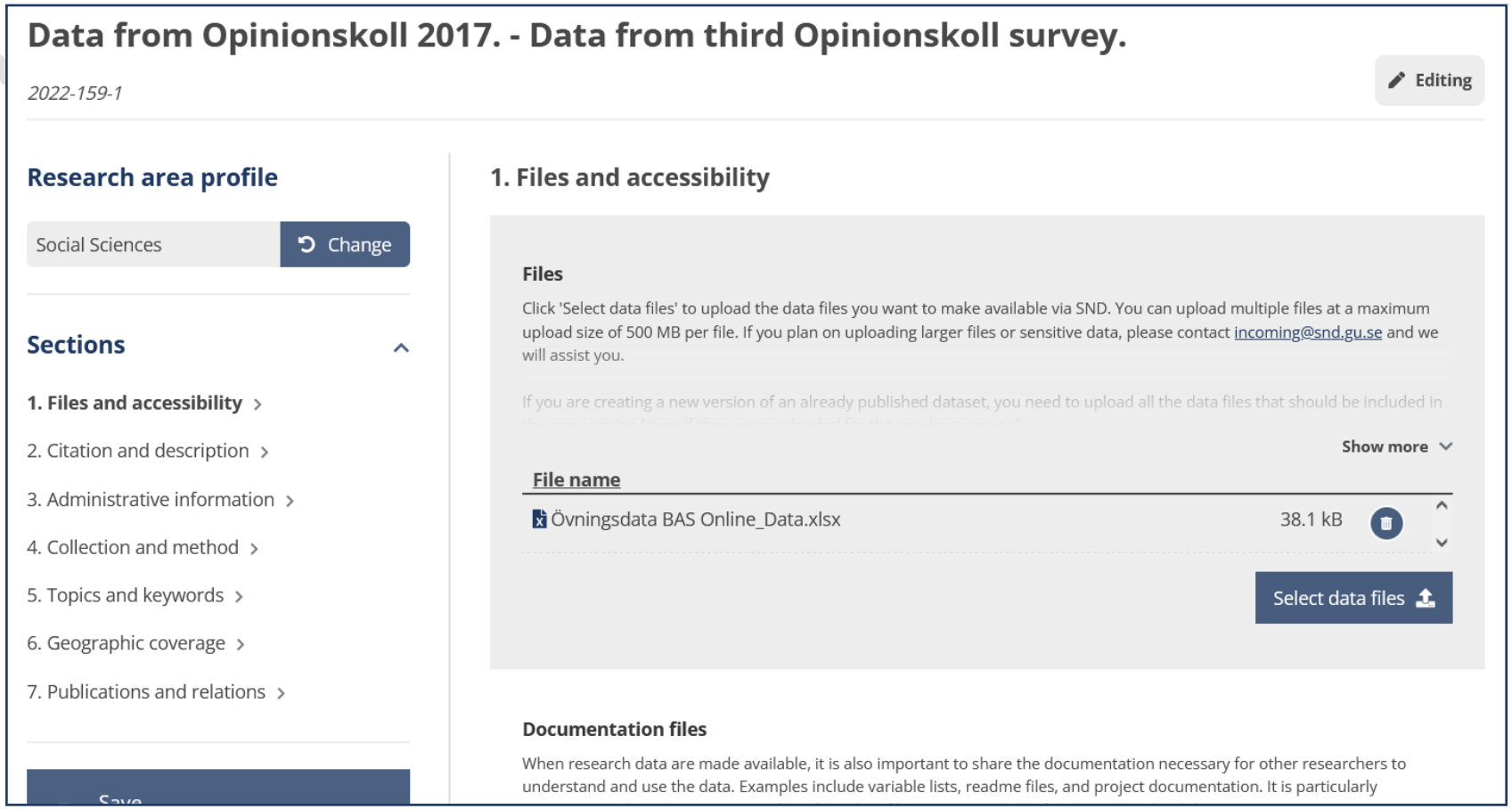


Research Data Catalogue

Search and find da

DORIS

Share data via the SND catalogue

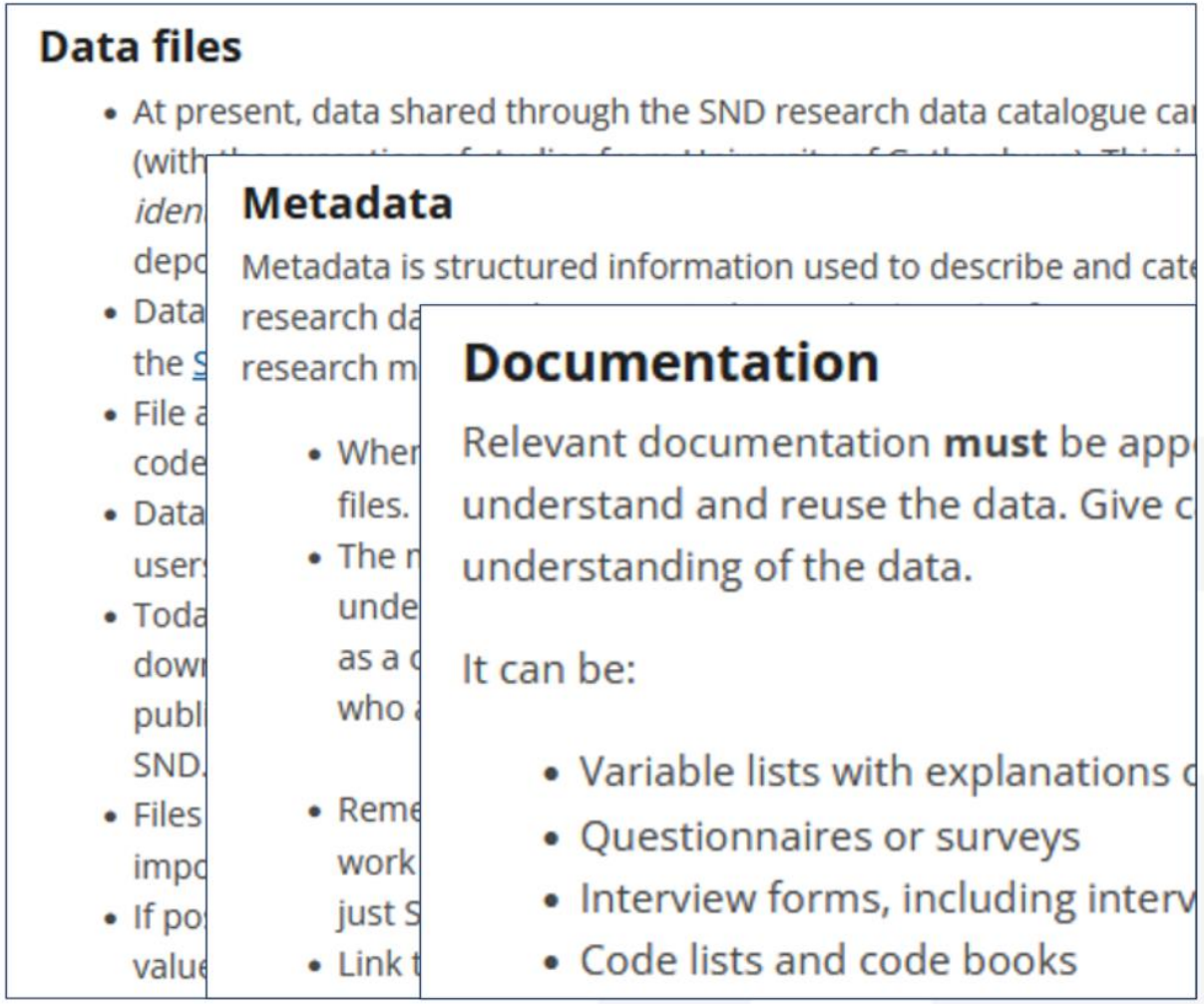


Research Data Catalogue

Search and find data

Share data via the SND catalogue

Share Data: Step by Step



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