The UiT research data web portal

[ui.no/forskningsdata](http://ui.no/forskningsdata)
The research data management life cycle and today’s seminar series

09:15 – 10:00
Search and cite research data

10:15 – 11:00
Select the appropriate license for your research data

11:15 – 12:00
Structure and document your research data

13:15 – 14:00
Share your research data

14:15 – 15:00
Write a data management plan*

*Despite the position of the planning phase (DMP) before structuring and sharing in the cycle, we leave DMP to the end, as it is wise to know about the various stages before you start planning.
Searching and citing research data

Glennda Villaflor
University Library

17.04.2018
Learning objectives

Part 1
• Understand the purpose of searching research data.
• Be familiar with 3 ways of searching data: the search engine, the archive, and the archive registry.

Part 2
• Understand the purpose of citing research data.
• Be familiar with the structure of the research data citation and the in-text citation.
## Some benefits of searching research data

**For the sharee**

- Get a deeper understanding of and an increased trust in the text you’re reading
- A source available for use in your own argumentation
- Avoid starting your data collection from scratch
- Adjust method and focus of your planned data collection

**For the sharer**

- Get better feedback on your research
- Make visible the full extent of your research output, including bi-directional referencing (article/dataset)
- Via metadata and repository guidelines, allow proper reuse and citation of data
- Increase citation of your research (Piwowar & Vision, 2013)
The research data management life cycle

Adapted original source: The University of California, Santa Cruz, Data Management LibGuide, Research Data Management Lifecycle, diagram, viewed May 2, 2016 at <http://guides.library.ucsc.edu/datamanagement>
Searching data: The search engine

DataCite

- Multi-disciplinary search engine
- Gathers metadata for each DOI assigned to an object
- Metadata free to access
- Combinatory searches possible
- [https://search.datacite.org/](https://search.datacite.org/)
Searching data: The search engine

TIP! If it says «restricted» next to a dataset, go to the repository and check whether you can get access through application

DataCite Search

- dyslexia children

7 Works

- Predicting Reading Performance from Pre-Literate Speech
Searching data: The search engine

Bielefeld Academic Search Engine (BASE)

• Multi-disciplinary search engine
• Indexes all resources providing an Open Archives Initiative interface
• Metadata free to access
• Refined search options
• https://www.base-search.net/
Tip! If it says «restricted» next to a dataset, go to the repository and check whether you can get access through application.

Searching data: The search engine

Your search
- dyslexi* child* doctype:7
- Boost open access documents

Linguistics tools
- Verbatim search
- Additional word forms
- Multilingual synonyms

Hit List

1. RADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia

- Title: RADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia
- Author: Ioannis Smyrnakis [claim]; Vassilios Andreadakis [claim]; Vassilios Selimis [claim]; Michail Kalaitzakis [claim]; Theodora Bachourou [claim]; Georgios Kaloutsakis [claim]; George D. Kymionis [claim]; Stelios Smirnakis [claim]; Ioannis M. Aslanides [claim]
- Description: Abstract: We introduce a novel, fast, objective, non-invasive method, named Rapid Assessment of Disabilities and Abnormalities in Reading (RADAR), that screens for features associated with the aberrant visual scanning of reading text, seen in dyslexia. Eye tracking shows Stelios Smirnakis? Then claim this publication in BASE/ORCID profile...
- Publisher: Figshare
- Year of Publication: 2017
- Document Type: Collection; Fileset
- Subjects: 80602 Computer-Human Interaction; 111704 Community Child Health; 111303 Vision Science
- Rights: CC-0; https://creativecommons.org/publicdomain/zero/1.0/
- Terms of Re-use: CCO
- URL: https://doi.org/10.6084/m9.figshare.4986365
Searching data:
The archive

UiT Open Research Data

- Multidisciplinary repository
- Metadata and data freely accessible
- Refined search options
- https://opendata.uit.no

UiT Open Research Data
support transparent and reproducible research
Searching data: The archive

**Zenodo**
- Multidisciplinary, open data
- Refined search options
- Operator AND, double quotation marks for phrases
- [https://zenodo.org/](https://zenodo.org/)

**Figshare**
- Multidisciplinary, open data
- How to use Figshare: [https://docs.figshare.com/](https://docs.figshare.com/)
- Refined search functions
- [https://figshare.com/](https://figshare.com/)
Searching data: The archive registry

Registry of Research Data Repositories (re3data)

Searchable registry for research data archives with open metadata

www.re3data.org
Searching data:
The archive registry

Browse by subject

Browse by subject, type of data, country

http://www.re3data.org/
Searching data: The archive registry

Main metadata very visible: subject, content, country, short description

http://www.re3data.org/
Searching data:
The archive registry

Information about access to data and link to guidelines

Repository details
Nord-Trøndelag Health Study

<table>
<thead>
<tr>
<th>General</th>
<th>Institutions</th>
<th>Terms</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of repository</td>
<td>Nord-Trøndelag Health Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional name(s)</td>
<td>Helseundersøkelsen i Nord-Trøndelag HUNT Study HUNT Databank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repository URL</td>
<td><a href="http://www.ntnu.edu/hunt">http://www.ntnu.edu/hunt</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject(s)</td>
<td>Social Sciences, Public Health, Health Services, Humanities and Social Sciences, Medicine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>The Nord-Trøndelag Health Study (The HUNT Study) is a unique database of personal health information from extensive studies. The fundamental strategy is that all individuals in the county will be invited to participate. The HUNT studies have been a successful and have resulted in extraordinary work in and with as is necessary for a successful health service. The data are collected in the HUNT surveys and are further health surveys in the county and account for a database with information about approximately 100,000 individuals and can be linked to national health registers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact</td>
<td><a href="http://www.ntnu.edu/contact">http://www.ntnu.edu/contact</a> <a href="mailto:hunt@medisin.ntnu.no">hunt@medisin.ntnu.no</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Repository details
Nord-Trøndelag Health Study

Policies (2)
Policy Name | Guidelines for administration and use of research data
URL | http://www.ntnu.edu/hunt/guidelines

Policy Name | Data access information
URL | http://www.ntnu.edu/hunt/data

Database access
Type of access to research data repository | open

Data access (3)
Type of access to data | open
Type of access to data | restricted
Data access restriction type(s) | other
Type of access to data | restricted
Data access restriction type(s) | feeRequired
Data licenses (1)
DataLicense | other
URL | http://www.ntnu.edu/hunt/guidelines
Searching data: The archive registry

Information about persistent identifiers, software, version control, etc.
Norwegian Centre for Research Data: administrative data

Individdata
- Norske spørreundersøkelser
- NSDs meningsmålingsarkiv
- European Social Survey
- Eurobarometer
- Skolevalg
- International Social Survey Programme (ISSP)
- Ungdom og historie
- Velferdssdata
- Generasjonsdatabasen
- Folketellingsdatabanken

Regionaldata
- Kommunedatabasen
- European Election Database
- NSDs Kirkedatabase
- The MacroDataGuide
- Kretsdata
- Nordiske data

Institusjonsdata
- Utdanningsstatistikk (DBH)
- Kirkedatabasen
- Data fra Innovasjon Norge
- Skolevalg

Data om det politiske system
- Storting
- Regjering
- Politiske parti
- Forvaltning
- Andre arkiv

Finn data med vår nye søkeportal (beta) >>
(En del datakilder er fremdeles ikke lagt inn)
Norwegian archives and other services for research data management

Forskningsrådet

Tilgjengeliggjøring av forskningsdata og data for forskning

Kunnskapsgrunnlag til KDs arbeid med en nasjonal strategi for tilgjengeliggjøring av data

https://www.forskningsradet.no/no/Artikkel/Apen_tilgang_til_forskningsdata/1254001013535
Register over åpne datasett i Norge

Her finner du publiserte datasett og åpne API-er fra offentlig sektor i Norge. Les mer om data.norge.no

Søk i datasett

FINN VIA TEMA:

Forvaltning og offentlig sektor (320)

Utdanning, kultur og sport (134)

Miljø (117)

Befolkning og samfunn (104)

Regioner og byer (94)

data.norge.no

www.europeandataportal.eu
Exercise

• Go to BASE (https://www.base-search.net/) and search data on a subject relevant to your work
• Go to re3data (http://www.re3data.org/) and find an archive relevant to your work

Questions to keep in mind
• Are you satisfied with the quantity and/or the quality of your results?
• Would you know how to refine or modify your search?
• Where do you go to get access to the data?
• How can you evaluate the relevance and quality of the data?

• Other questions we should ask ourselves when searching?
Citing research data

Hansen, Pemille, 2016, "Replication data for: What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development", doi:10.18710/JEWIVV.
Citing research data

Notes

1. There are three Norwegian CHILDES corpora, but the third is transcribed in a Trondheim dialect, and could not be processed by the automatic tagger.

2. The complete frequency list is available at http://dx.doi.org/10.18710/JEWIVW together with data on AoA, VSoA, word class and frequency for the CDI words analysed here (Hansen, 2016).


Hansen, P. (2016). Replication data for: What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development (UiT Open Research Data Dataverse, V1). doi:10.18710/JEWIVW

Hansen, P., Simonsen, H. G., Łuniewska, M., & Haman, E. (in press). Validating the psycho-
The purpose of the research data citation

Facilitate description and information Retrieval
(author, archive, subset ...)

Facilitate access and persistence
(reference in article, digital, url/pid)

Facilitate verification and reproducibility
(support linkage data-claim)

Facilitate integration
(data as part of the scholarly research and publishing ecosystem)

(Altman & Crosas, 2013, p. 64)
## Structure of reference: Obligatory elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>Unique string that identifies the dataset (doi, handle)</td>
</tr>
<tr>
<td>Author</td>
<td>The researcher(s) having produced the data and are authors of the corresponding journal article</td>
</tr>
<tr>
<td>Title</td>
<td>Name of the dataset</td>
</tr>
<tr>
<td>Publisher</td>
<td>Name of the archive</td>
</tr>
<tr>
<td>Year of publication</td>
<td>Moment when the data are made available (in the case of embargo, use moment when the embargo period is over)</td>
</tr>
</tbody>
</table>

(Altman & Crosas, 2013; Starr & Gastl, 2011)
### Structure of reference: Recommended elements

<table>
<thead>
<tr>
<th>Version</th>
<th>If dataset changes, the version number changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of data</td>
<td>e.g. dataset, corpus, picture archive</td>
</tr>
<tr>
<td>Related identifier</td>
<td>Full dataset in the case of subset use</td>
</tr>
</tbody>
</table>

(Altman & Crosas, 2013; Starr & Gastl, 2011)
Example


Bibliographic reference:
Coretta, Stefano, 2016, "Data on Icelandic pre-aspiration",
doi:10.18710/7NLJSG, UiT Open Research Data Dataverse, V1
In-text citation: Full dataset

Citing dataset:
Usual data citation rules

Citing a file in the dataset:
Use the file name
Example


Bibliographical reference:
Coretta, Stefano, 2016, "Data on Icelandic pre-aspiration", doi:10.18710/7NLJSG, UiT Open Research Data Dataverse, V1

In-text citation
This is just a nonsense text that is actually about consonants (Coretta, 2016). Now I will refer to a particular file in the dataset (Coretta, 2016, file name: brs02_20160531.wav), which I have looked at more closely because it contains data on the consonants [t,k], which is the subject of my own study.
In-text citation: Specific files

“Suggestion: [...] citing at the lowest level for which there is a DOI.” (Beaver & Dubinsky, 2017)

You can also cite specific files, as long as they have a persistent identifier.
Fill out our evaluation form!

bit.ly/ubevalen

Teacher’s name: Data group
Date: 17.04.2018
Course code or title of session/course: Research data

Thanks and good luck!

research-data@support.uit.no
References:


DataCite Metadata Schema: [https://schema.datacite.org/](https://schema.datacite.org/)