UiT Norges arktiske universitet

How to archive research data

Tromsø, 18 April 2023

University Library

Andreas Klein, PhD

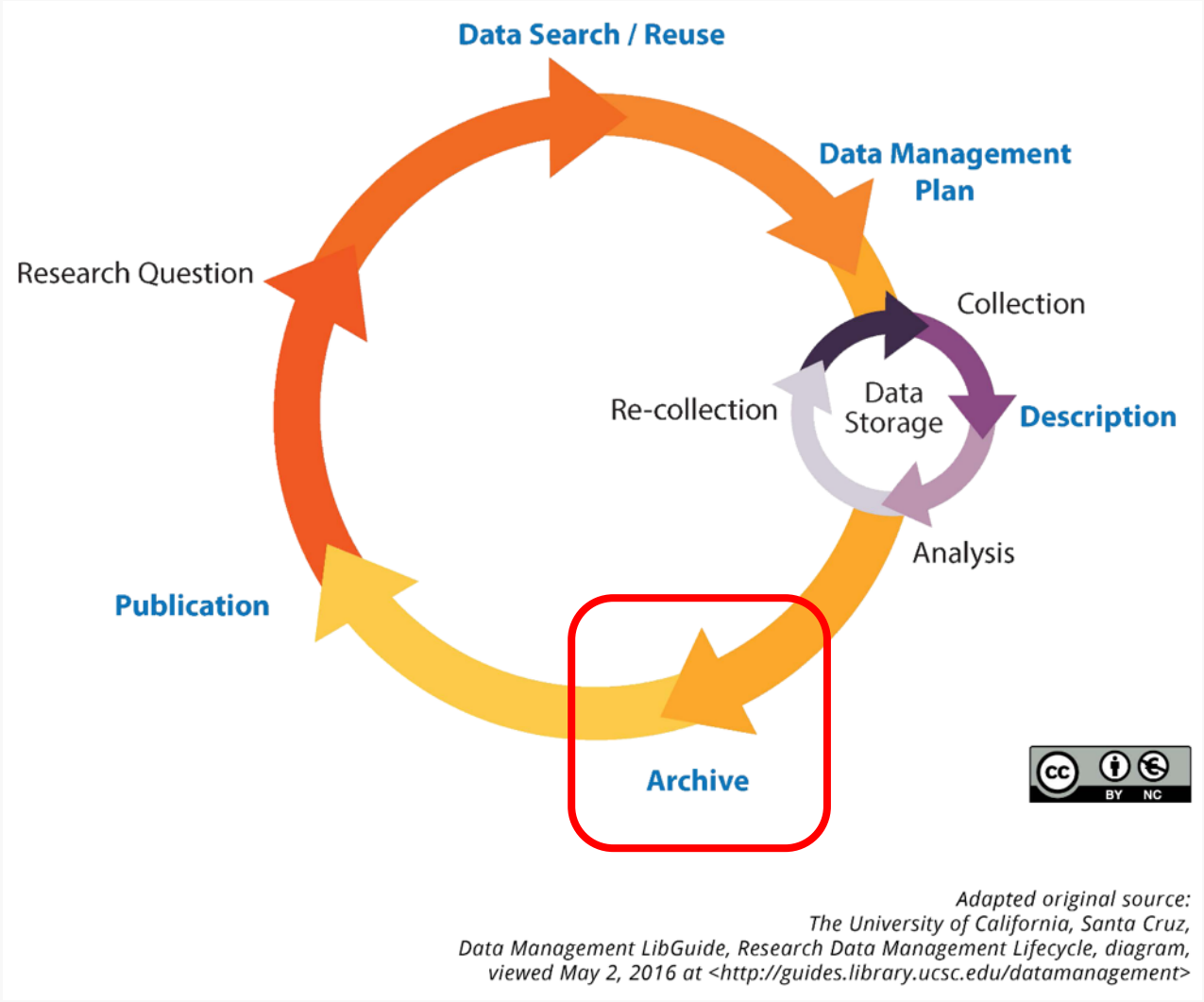
Anna Karpova, Academic Librarian

Research data @ UiT CC BY 4.0

Learning objectives

- Understand the purpose of archiving (and sharing of) data and/or metadata.
- Be familiar with important criteria in choosing a data repository.
- Know how to proceed when searching for suitable data repositories.
- Know what to do in case your research data contain sensitive information.

CC BY



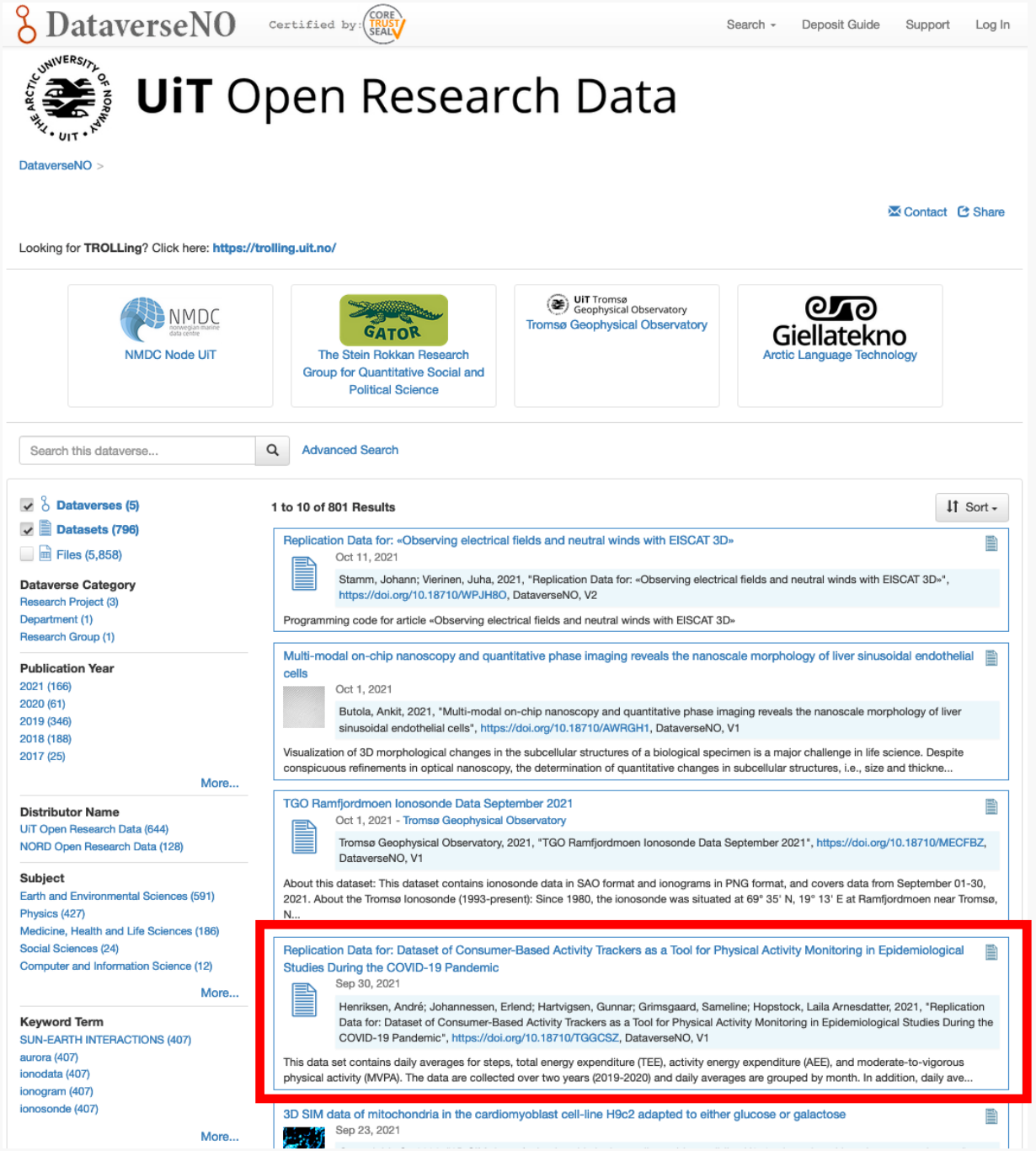
Adapted original source:
The University of California, Santa Cruz,
Data Management LibGuide, Research Data Management Lifecycle, diagram,
viewed May 2, 2016 at <http://guides.library.ucsc.edu/datamanagement>

Archived datasets

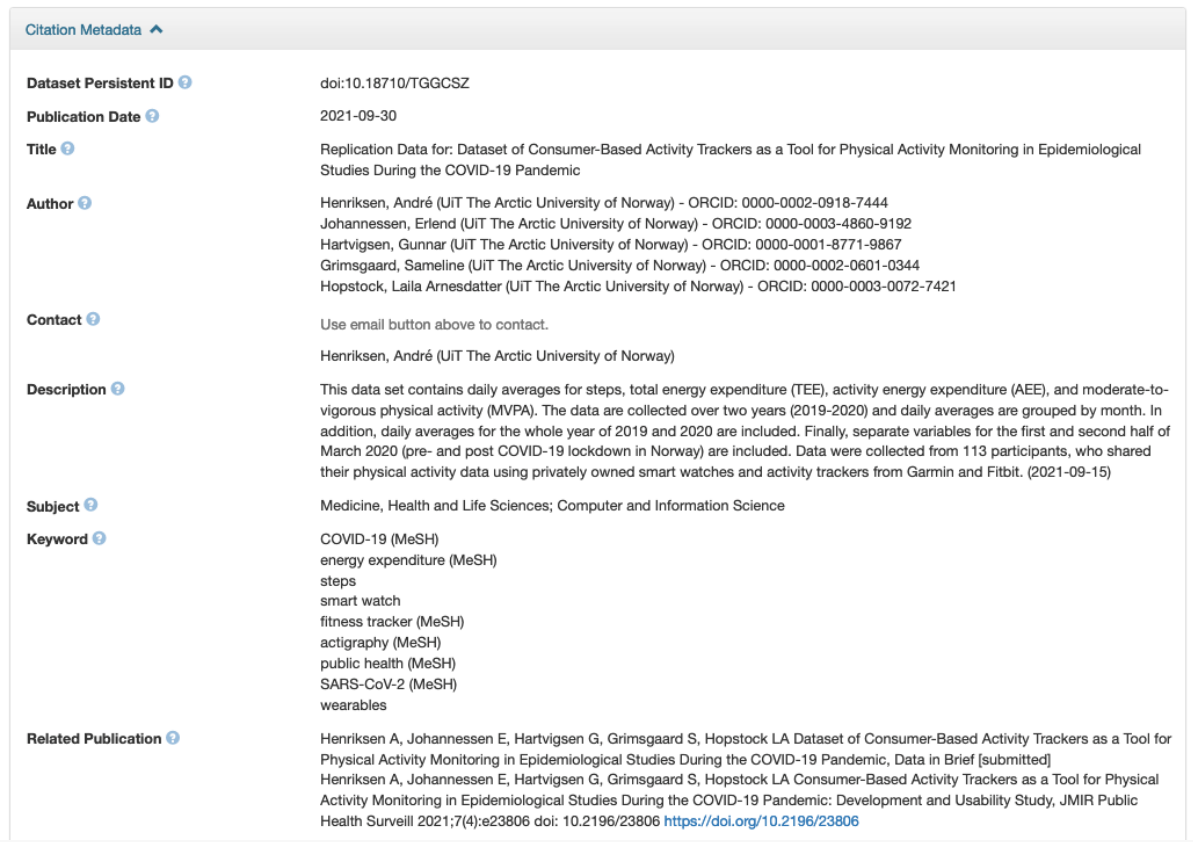
CC BY

Metadata:

Data about data

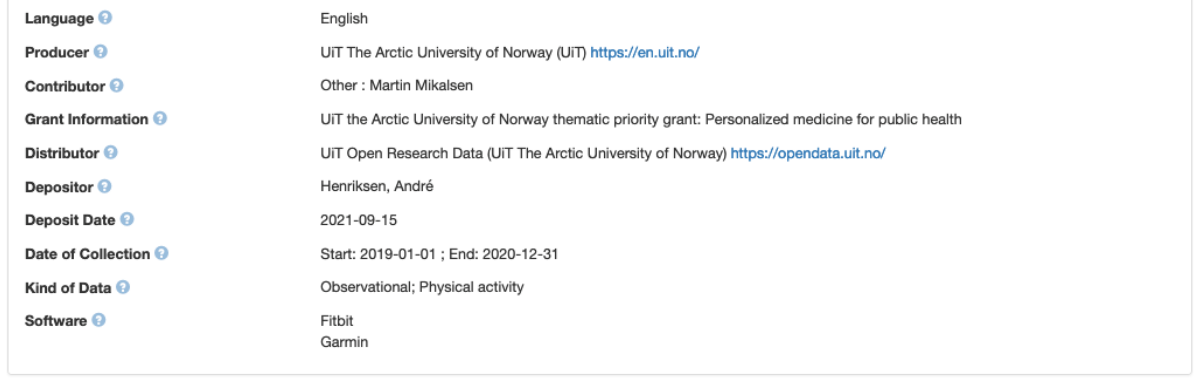


CC BY



Citation Metadata

Dataset Persistent ID	doi:10.18710/TGGCSZ
Publication Date	2021-09-30
Title	Replication Data for: Dataset of Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic
Author	Henriksen, André (UiT The Arctic University of Norway) - ORCID: 0000-0002-0918-7444 Johannessen, Erlend (UiT The Arctic University of Norway) - ORCID: 0000-0003-4860-9192 Hartvigsen, Gunnar (UiT The Arctic University of Norway) - ORCID: 0000-0001-8771-9867 Grimsgaard, Sameline (UiT The Arctic University of Norway) - ORCID: 0000-0002-0601-0344 Hopstock, Laila Arnesdatter (UiT The Arctic University of Norway) - ORCID: 0000-0003-0072-7421
Contact	Use email button above to contact. Henriksen, André (UiT The Arctic University of Norway)
Description	This data set contains daily averages for steps, total energy expenditure (TEE), activity energy expenditure (AEE), and moderate-to-vigorous physical activity (MVPA). The data are collected over two years (2019-2020) and daily averages are grouped by month. In addition, daily averages for the whole year of 2019 and 2020 are included. Finally, separate variables for the first and second half of March 2020 (pre- and post COVID-19 lockdown in Norway) are included. Data were collected from 113 participants, who shared their physical activity data using privately owned smart watches and activity trackers from Garmin and Fitbit. (2021-09-15)
Subject	Medicine, Health and Life Sciences; Computer and Information Science
Keyword	COVID-19 (MeSH) energy expenditure (MeSH) steps smart watch fitness tracker (MeSH) actigraphy (MeSH) public health (MeSH) SARS-CoV-2 (MeSH) wearables
Related Publication	Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA Dataset of Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic, Data in Brief [submitted] Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic: Development and Usability Study, JMIR Public Health Surveill 2021;7(4):e23806 doi: 10.2196/23806 https://doi.org/10.2196/23806

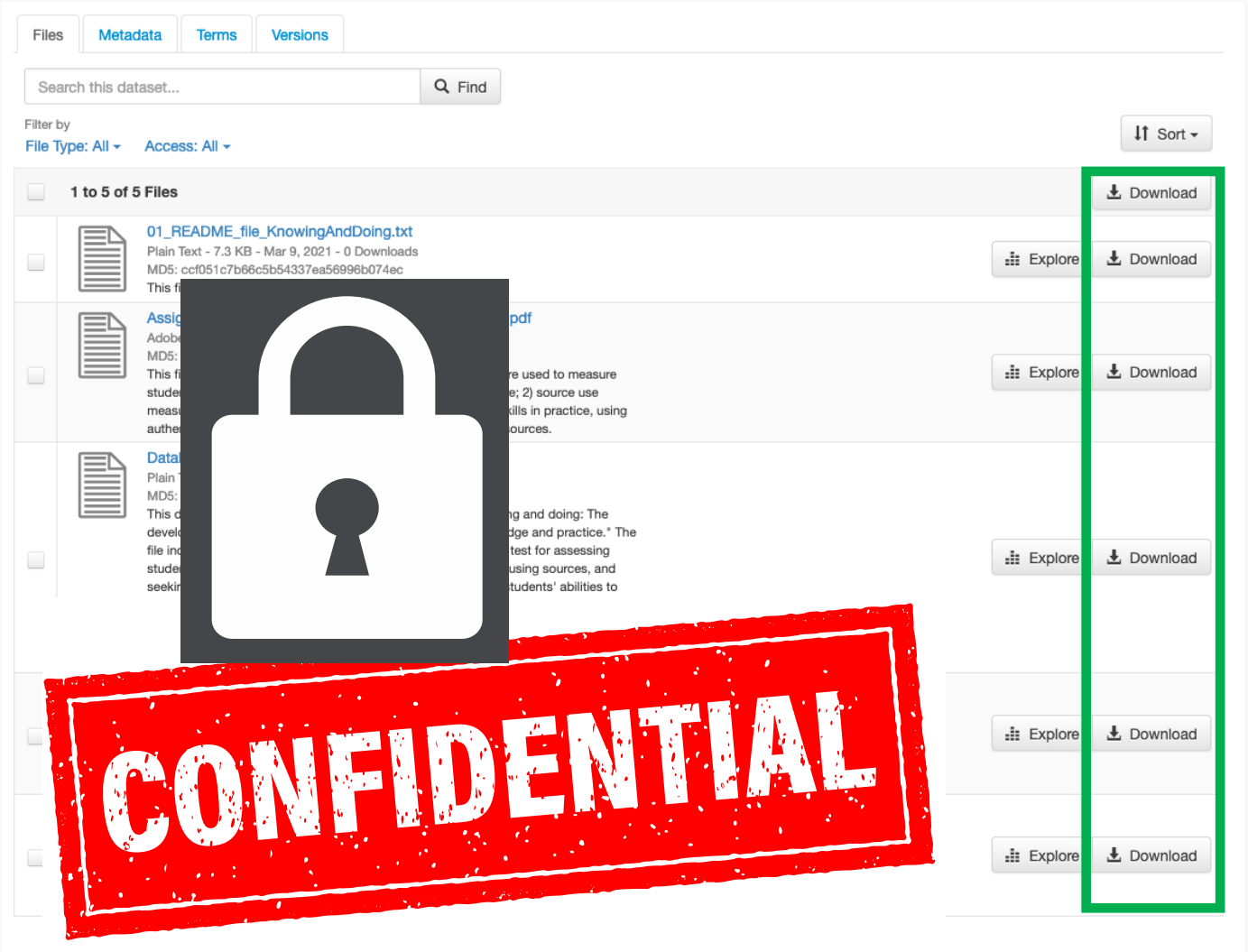


Language	English
Producer	UiT The Arctic University of Norway (UiT) https://en.uit.no/
Contributor	Other : Martin Mikalsen
Grant Information	UiT the Arctic University of Norway thematic priority grant: Personalized medicine for public health
Distributor	UiT Open Research Data (UiT The Arctic University of Norway) https://opendata.uit.no/
Depositor	Henriksen, André
Deposit Date	2021-09-15
Date of Collection	Start: 2019-01-01 ; End: 2020-12-31
Kind of Data	Observational; Physical activity
Software	Fitbit Garmin

Metadata: Data about data

CC BY

Files: Archived with open or restricted access



CC BY

Why archive data?

Transparency

Reproducibility/Replicability

Encourage improvement & validation of research methods

Fulfill data sharing requirements

Ensure future access to the data

Make your data *FAIR*

Increase the impact & visibility of your research

Enable data reuse for research & innovation
Get credit
Find new collaborations
Comply with expectations during evaluation

Requirements and expectations

EU, Horizon 2020

As open as possible, as closed as necessary

H2020 Programme: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

EU, Horizon Europe (2021-2027)
... robust exceptions to this rule, where access to data needs to be protected and Intellectual Property Rights protected.

Research Council of Norway

Open by default

Open Access to Research Data

Norwegian Ministry of Education and Research

Basic principle 1: Research data must be as open as possible, as closed as necessary.

Basic principle 2: Research data should be managed and curated to take full advantage of their potential.

Basic principle 3: Decisions concerning archiving and management of research data must be taken within the research community

National strategy on access to and sharing of research data

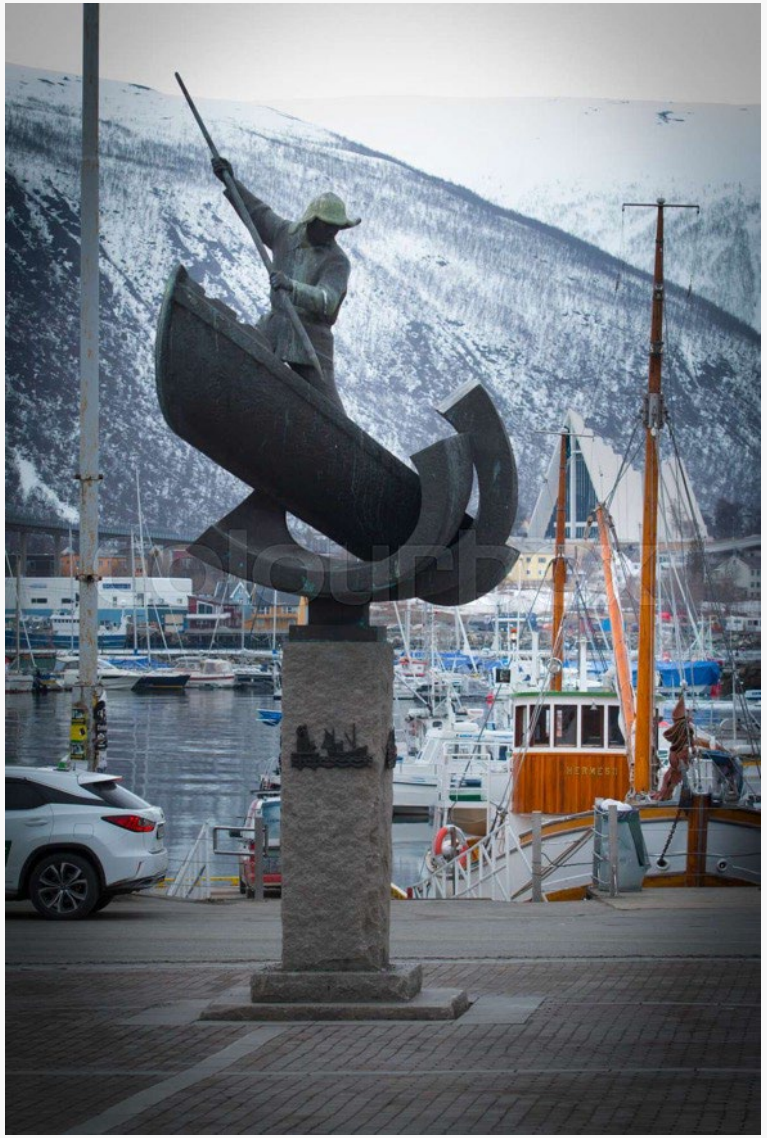
CC BY

Requirements and expectations

UiT The Arctic University of Norway

The researcher shall make research data openly available for further use to all relevant users, providing there are no legal, ethical, security or commercial reasons for not doing so.

Principles and guidelines for management of research data at UiT (updated 2021)



CC BY

Requirements and expectations: Journals

Data Sharing and Materials Availability

After publication, all data and materials (including computer codes) necessary to understand, assess and extend the conclusions of the manuscript must be available to any reader of the Science Partner Journals. All reasonable requests for data or materials must be fulfilled. Unreasonable restrictions on data or material availability may prevent publication.

http://www.sciencemag.org/authors/science-journals-editorial-policies

CC BY

What, when and where to archive?

What

The research communities are responsible for deciding what is worth archiving/sharing.

Using data from a third party: Reuse license and contract decide

When

As early as possible!

- Scientific publications: No later than at the time of publication.
- Other data: Normally when the project ends, or after an embargo period.

Where

“Rights to use and publish research data shall not be transferred to commercial actors without UiT retaining the rights to make the data available for reuse in teaching and research.”

- Reliable archives where UiT is ensured access to the data
- Funders and journals might point to specific repositories

CC BY

How to choose a repository

1. Use an external data archive or repository already established for your research domain to preserve the data according to recognised standards in your discipline.

2. If available, use an institutional research data repository, or your research group's established data management facilities.

3. Search for other data repositories in: re3data.org.

Based on https://www.openaire.eu/opendatapilot-repository-guide

"Orange County Archives, 1980s" by Orange County Archives is licensed under CC BY 2.0

How to choose a repository

Questions you should ask, regardless of whether you plan to archive openly or whether only (parts of) metadata can be shared.

1. Is the data repository reputable?
2. Will the repository take the data you want to deposit?
3. Will the data be safe in legal terms?
4. Will the archive sustain the value of the data?
5. Will the archive support analysis and track data usage?

(Whyte, 2015)

"Orange County Archives, 1980s" by Orange County Archives is licensed under CC BY 2.0

UiT Open Research Data

- Archiving, reuse and citation of open research data.

- Upload: Employees and students at UiT (Feide login).
- Download and reuse: Everyone.

- Built on the international platform Dataverse.
- CoreTrustSeal certified

- Datasets curated by the University Library

- Link to the repository: opendata.uit.no
- Deposit guidelines can be found at info.dataverse.no

CC BY

Attend the webinar later today (12:15–14:00), to learn more about the repository!

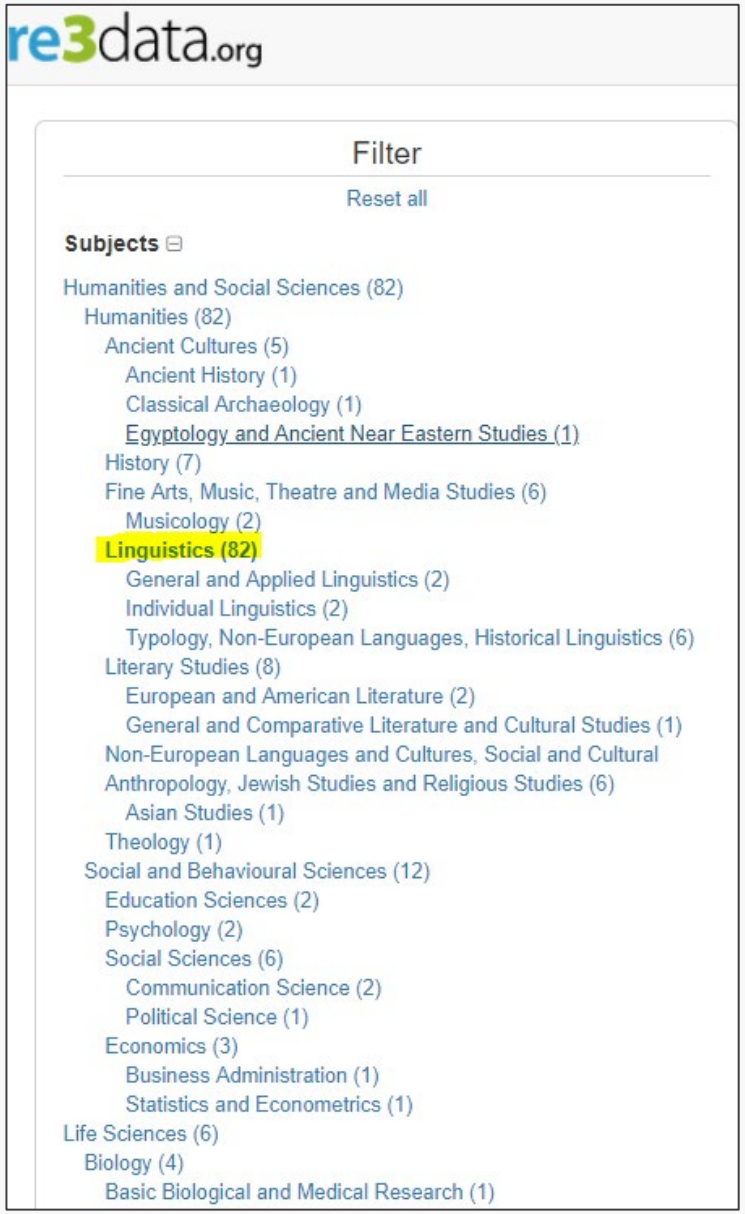
Find Zoom link at uit.no/forskningsdata

CC BY

Finding a repository for your data

- About 3100 different repositories (2023)
- Multidisciplinary, search engine
- A lot of information about the repositories: terms of use, standards, persistent identifiers, software, version control, etc.
- http://www.re3data.org/

Browse by discipline



Browse by discipline

CLARIN-ERIC

Common Language Resources and Technology Infrastructure - European Research Infrastructure Consortium

Subject(s)	Humanities and Social Sciences · Linguistics · Artificial Intelligence, Image and Language Processing · Humanities · Computer Science · Computer Science, Electrical and System Engineering · Engineering Sciences
Content type(s)	Standard office documents · Audiovisual data · Scientific and statistical data formats · Raw data · Plain text
Country	Netherlands · European Union

CLARIN is a European Research Infrastructure for the Humanities and Social Sciences, focusing on language resources (data and tools). It is being implemented and constantly improved at leading institutions in a large and growing number of European countries, aiming at improving Europe's multi-linguality competence. CLARIN provides several services, such as access to language data and tools to analyze data, and offers to deposit research data, as well as direct access to knowledge about relevant topics in relation to (research on and with) language resources. The main tool is the 'Virtual Language Observatory' providing metadata and access to the different national CLARIN centers and their data.

Pacific and Regional Archive for Digital Sources in Endangered Cultures

PARADISEC

Subject(s)	Linguistics · Typology, Non-European Languages, Historical Linguistics · Humanities · Humanities and Social Sciences · Musicology · Fine Arts, Music, Theatre and Media Studies
Content type(s)	Standard office documents · Images · Scientific and statistical data formats · Audiovisual data
Country	Australia

PARADISEC (the Pacific And Regional Archive for Digital Sources in Endangered Cultures) offers a facility for digital conservation and access to endangered materials from all over the world. Our research group has developed models to ensure that the archive can provide access to interested communities, and conforms with emerging international standards for digital archiving. We have established a framework for accessioning, cataloguing and digitising audio, text and visual material, and preserving digital copies. The primary focus of this initial stage is safe preservation of material that would otherwise be lost, especially field tapes from the 1950s and 1960s.

OLAC

Open Language Archives Community

Subject(s)	Linguistics · Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Studies and Religious Studies · Artificial Intelligence, Image and Language Processing · Humanities · Humanities and Social Sciences · Computer Science

List with main information, including icons that indicate access, persistent identifier, use of standards, etc.

Archiving data with restricted access

UiT

"The researcher shall make research data openly available for further use to all relevant users, providing there are no **legal, ethical, security or commercial** reasons for not doing so."

(Principles and guidelines for management of research data at UiT).

Eventually, support services for all types of data shall be in place*

* work in progress

CC BY

Which data are you allowed to archive openly?

Archive with
- open access to primary data, processed data and metadata

Archive with
- closed/restricted access to primary data
- open access to processed data and metadata

Archive with
- closed/restricted access to primary data and processed data
- open access to metadata

Archive with
- closed/restricted access to primary data, processed data and metadata

Open (full access) ⟷ **Closed (no access)**

The degree of sensitivity and what is authorised decides where data can be stored, and whether/where the data can be archived.

CC BY

Archiving data with restricted access

UiT

“The researcher shall make research data openly available for further use to all relevant users, providing there are no legal, ethical, security or commercial reasons for not doing so.”

Norwegian Centre for Research Data (NSD)

Datasets published on NSD with data are freely available for download. Datasets published without data are available **pending application approval**, and in some cases, only following authorisation from researcher/producer. Specific rules apply to personal data.

Not sure where to archive your data?
Contact us at researchdata@hjelp.uit.no

CC BY

Are you sitting on «old» data?

- Do you have previously collected (and processed) data that you want to archive and share?

- Contact researchdata@hjelp.uit.no for tips on good and efficient data management.

"World's Messiest Office Cubicle Discovered in Colorado" by Jeffrey Beall is licensed under CC BY-ND 2.0

More information and help

UiT Research Data Portal: https://uit.no/forskningsdata/

Email: researchdata@hjelp.uit.no

CC BY

References

European Commission. (2017). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020.* Version 3.2. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Kunnskapsdepartementet. (2017). *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata.* https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf

Norges forskningsråd. (2017). *Tilgjengeliggjøring av forskningsdata* (revidert 2017). Hentet fra www.forskningsradet.no/publikasjoner

UiT Norges arktiske universitet. (2021). *Prinsipper og retningslinjer for forvaltning av forskningsdata ved UiT.* https://uit.no/regelverk/sentraleregler#v-pills-742423

Whyte, A. (2015). *Where to keep research data. DCC checklist for evalutating data repositories*. V1. Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data

CC BY

researchdata@hjelp.uit.no

Andreas Klein

Anna Karpova

 CC BY 4.0