Research Data Management and Open Science at NTNU

Ingrid Heggland, PhD
NTNU University Library, Research Data @NTNU

ELIXIR DMP workshop 15-16 June 2021

OPEN SCIENCE NTNU

Research Data @NTNU

- Central support service for research data
  - Coordinated by the library in close collaboration with IT
  - Support for data management troughout the data lifecycle

https://innsida.ntnu.no/researchdata

research-data@ntnu.no

OPEN SCIENCE NTNU

What is Open Science?

- "to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction"

(OECD 2015:7, http://dx.doi.org/10.1787/5jrs2f963zs1-en).

***- "extending the principles of openness to the whole research cycle**, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done"*

(https://www.fosteropenscience.eu/content/what-open-science-introduction).

NTNU | Norwegian University of Science and Technology

RCN 2020: New policy for Open Research

"Open research means scientific practice where processes and results are openly available under conditions that promote quality and knowledge development, including the sharing and use of the research-based knowledge in a socially responsible way" (Research Council of Norway 2020).

OPEN SCIENCE NTNU

Open and free sharing of knowledge in a digital world

Responsible, reproducible and accessible science

«As open as possible, as closed as necessary»

NTNU | Norwegian University of Science and Technology

OPEN SCIENCE NTNU

- Collaboration
- Dissemination
- Transparency
- Reproducibility

NTNU | Norwegian University of Science and Technology

Open Science at NTNU

- New Open Science policy from 2021
  - Open Access (publications) from 2014
  - (Open) research data from 2018

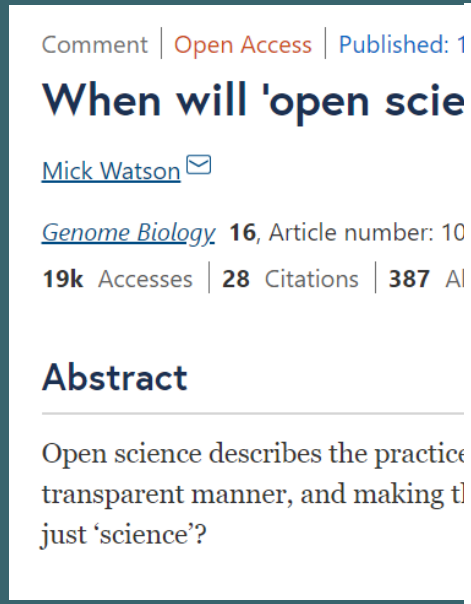
- NTNU University Library point of contact and coordination of Open Science
  - Support, tools and guidance
  - Collaboration with other departments, faculties and institutes

OPEN SCIENCE NTNU

- Open Access
- Open Data
- Open Source
- Citizen Science
- Open Educational Resources

NTNU | Norwegian University of Science and Technology

Open Science: Just Science done right?



Comment | Open Access | Published: 1

When will 'open scie

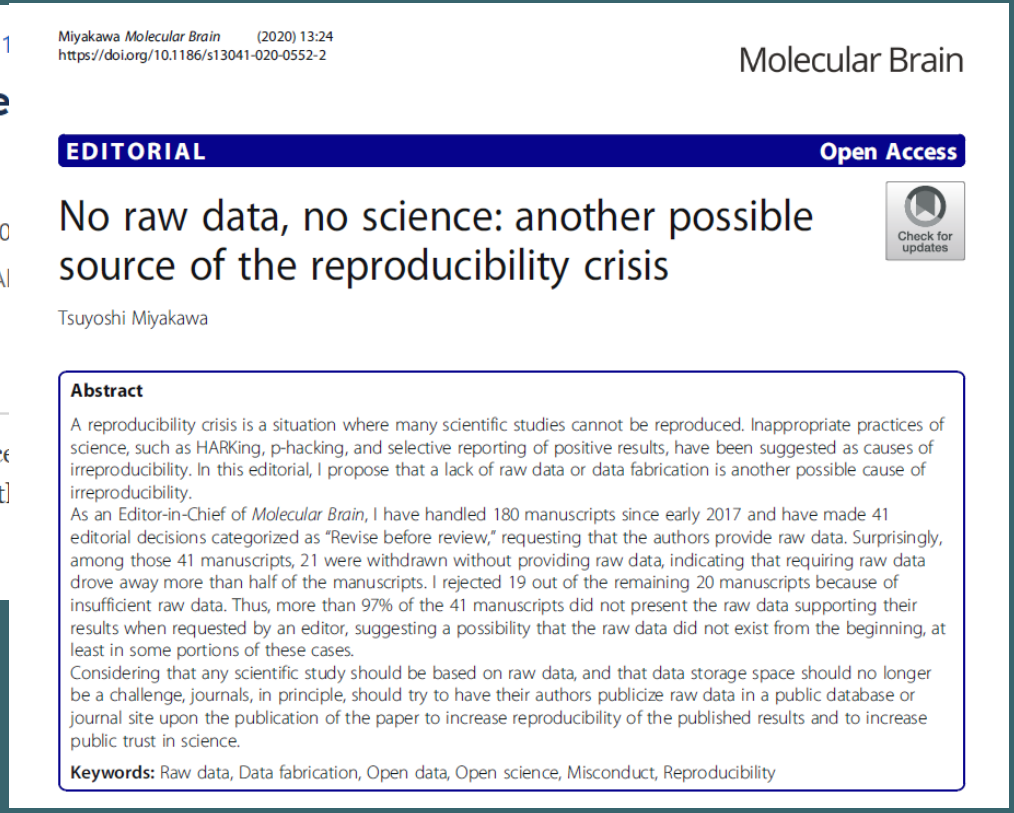
Mick Watson

Genome Biology **16**, Article number: 10

19k Accesses | **28** Citations | **387** Al

Abstract

Open science describes the practice transparent manner, and making t just 'science'?



Miyakawa *Molecular Brain* (2020) 13:24
https://doi.org/10.1186/s13041-020-0552-2

Molecular Brain

EDITORIAL Open Access

No raw data, no science: another possible source of the reproducibility crisis

Tsuyoshi Miyakawa

Abstract

A reproducibility crisis is a situation where many scientific studies cannot be reproduced. Inappropriate practices of science, such as HARKing, p-hacking, and selective reporting of positive results, have been suggested as causes of irreproducibility. In this editorial, I propose that a lack of raw data or data fabrication is another possible cause of irreproducibility.

As an Editor-in-Chief of *Molecular Brain*, I have handled 180 manuscripts since early 2017 and have made 41 editorial decisions categorized as "Revise before review," requesting that the authors provide raw data. Surprisingly, among those 41 manuscripts, 21 were withdrawn without providing raw data, indicating that requiring raw data drove away more than half of the manuscripts. I rejected 19 out of the remaining 20 manuscripts because of insufficient raw data. Thus, more than 97% of the 41 manuscripts did not present the raw data supporting their results when requested by an editor, suggesting a possibility that the raw data did not exist from the beginning, at least in some portions of these cases.

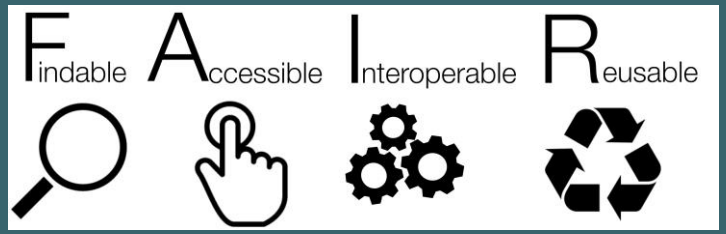
Considering that any scientific study should be based on raw data, and that data storage space should no longer be a challenge, journals, in principle, should try to have their authors publicize raw data in a public database or journal site upon the publication of the paper to increase reproducibility of the published results and to increase public trust in science.

Keywords: Raw data, Data fabrication, Open data, Open science, Misconduct, Reproducibility

OPEN SCIENCE NTNU

Research data at NTNU: policy

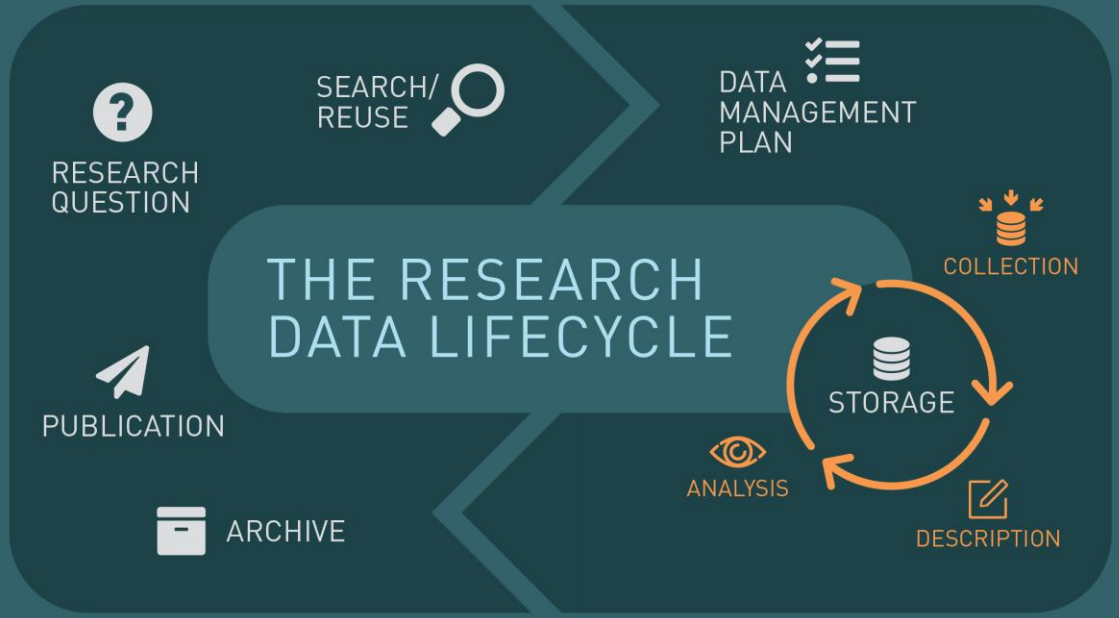
- **As open as possible, as closed as necessary**
- **Data should be:**



- All research projects should have a DMP
- NTNU guarantees for access to basic infrastructure and support
- The researcher is responsible for the quality and handling of data

Open Data

OPEN SCIENCE
NTNU



NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
NTNU

Open Science and Research Data Management go hand in hand

Research Data

FOR EMPLOYEES

What is Research Data @NTNU?

- a central support service for research data
- a service for researchers and students at NTNU
- a contact point for faculties and institutes at NTNU
- a collaboration between the University Library and NTNU IT

NTNU requires good research data management

Research data at NTNU should be managed according to best practice, and be as open as possible, as stated in the Policy and Plan of action for research data.

Data Management Plan (DMP)

All research projects at NTNU should develop a Data Management Plan describing how the research data will be managed.

NTNU Open Research Data

Research data at NTNU can be published and shared openly in our repository for research data, NTNU Open Research Data.

Data Management

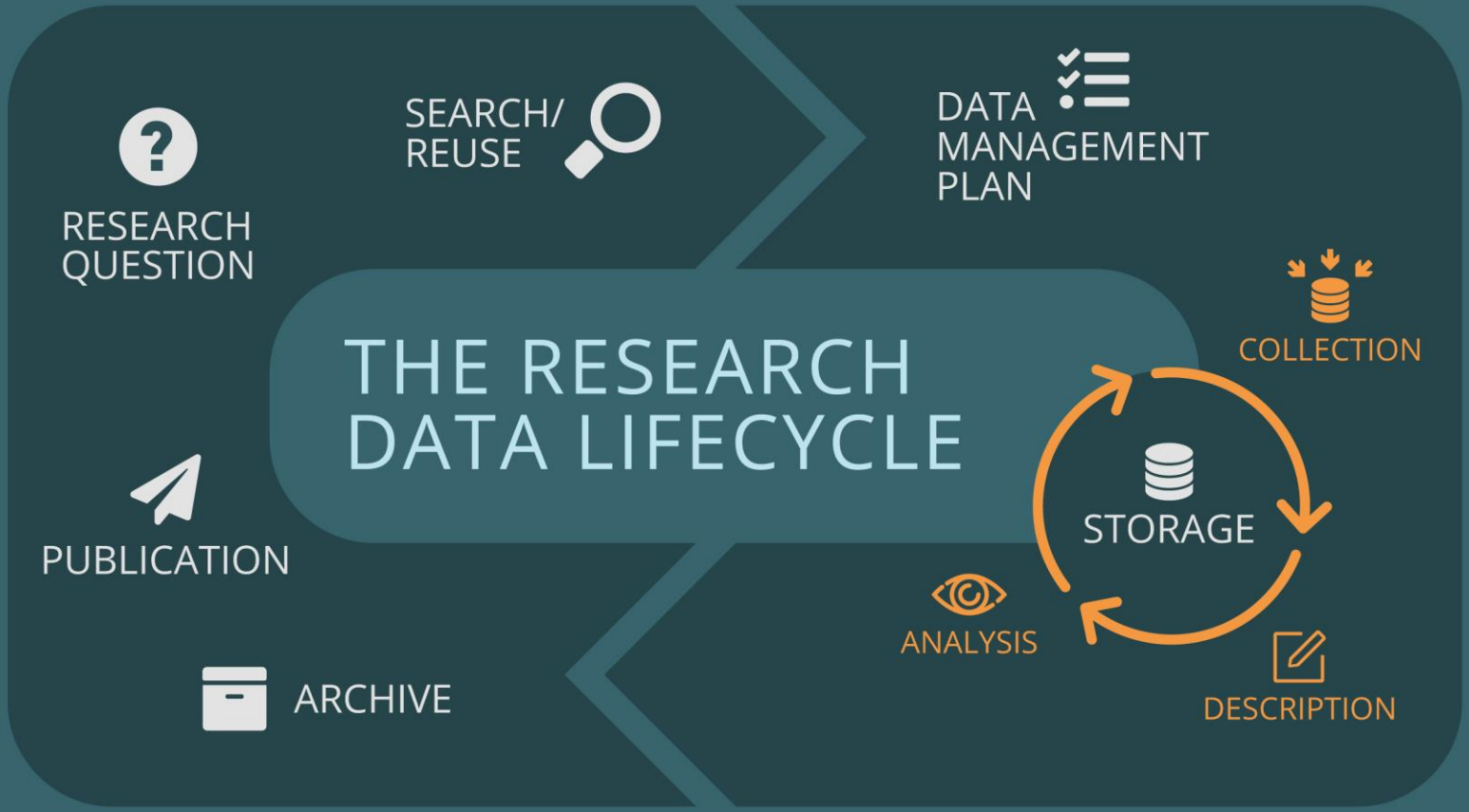
- Search for data
- Data Management Plan (DMP) and planning
- Storage and active management of research data
- Archiving and publishing of research data
- Training, guidance and support

Useful resources

- ELIXIR Norway - support and tools for life science research
- Course in data management on FOSTER e-learning platform (EU project)
- MOOC on Open Science from TUDelft
- PhD on Track
- Mantra Research Data Management Training
- MOOC on Research Data Management and Sharing from Coursera
- Cessda Data Management Expert Guide

Contact

If you can not find the information you are looking for, please send an email to research-data@ntnu.no

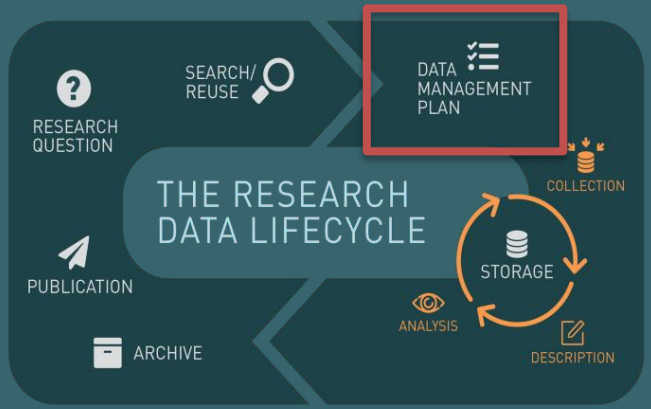


NTNU | Norwegian University of Science and Technology

Data Management Plan (DMP)

- A useful tool for research projects!
  - Planning data management during the project (and after)

- Both a formal and «living» document
  - Requirements from funders and NTNU
  - Revise and update = documentation



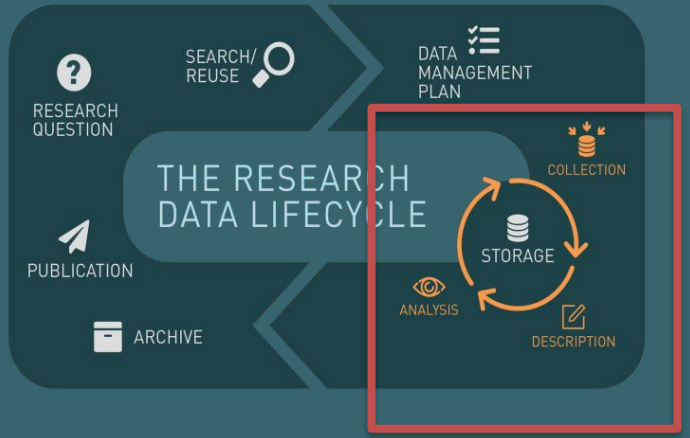
What should a DMP include?

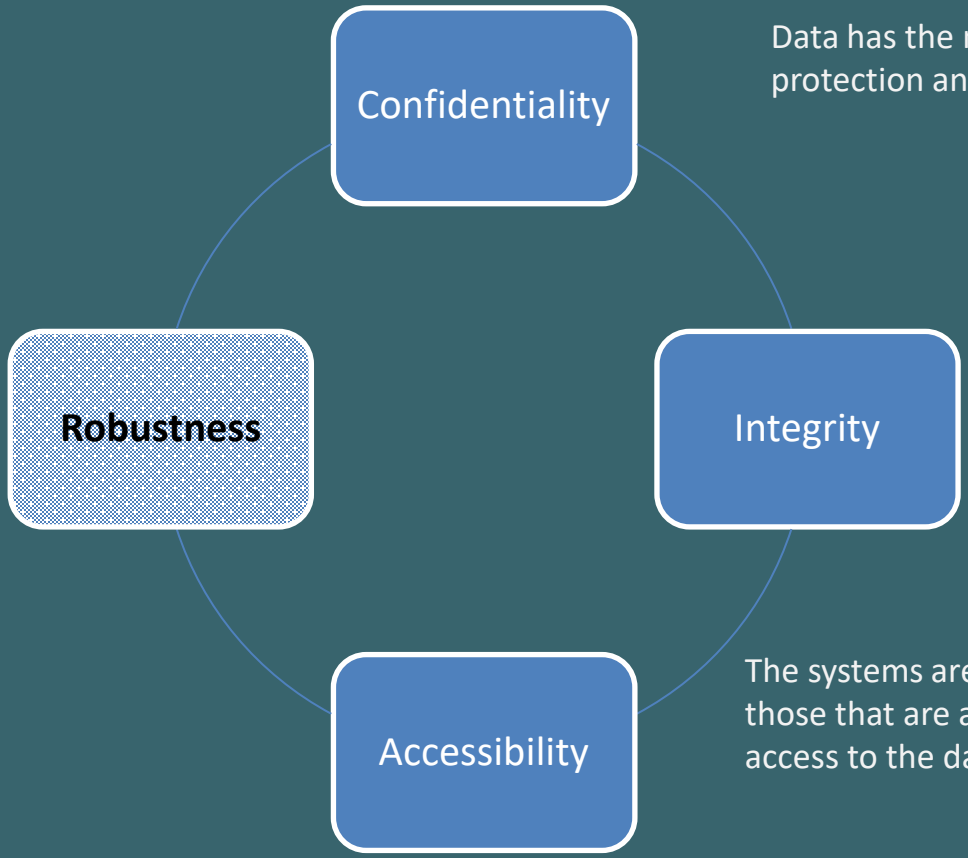
- Data collection, methods
- Description of data, formats, organising, metadata
- Storing, archiving, backup, sharing
- Intellectual rights, licenses, privacy, ethics
- Costs, responsibility

- Guidelines, tools and support at NTNU:
  - https://innsida.ntnu.no/wiki/-/wiki/English/Data+management+plan

Active storage of data

- Storage during the project period
  - Confidentiality, Integrity, Accessibility (CIA), data volume etc…
- Storage guide for NTNU:
  - https://innsida.ntnu.no/wiki/-/wiki/Norsk/Lagringsguide





Data has the necessary levels of protection and access controls.

The information should be updated, correct and complete.

The systems are up and running and those that are authorized,have access to the data.

Information security and sensitive data

- Sensitive (research) data
  - Data requiring protection: Confidential data
    - Personal information
    - Trade secrets
    - Commercial purposes, patents etc
    - Etc…

- Research collecting personal data:
  - Projects at NTNU must be reported to NSD
    - Exception: Healt research at the

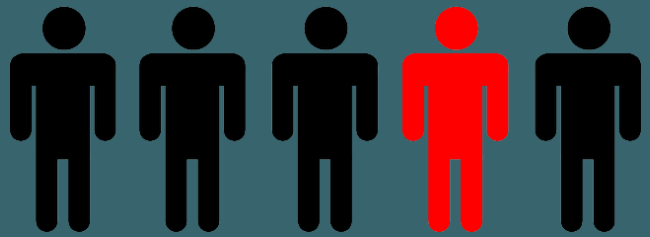
https://innsida.ntnu.no/wiki/-/wiki/English/Collection+of+personal+data+for+research+projects

WHAT IS PERSONAL DATA?

Any information that identifies a physical person, directly or indirectly

Examples:

- Name, ID-number, address, telephone number etc.
- IP-adress, location information
- Also images, sound, video, email, voice

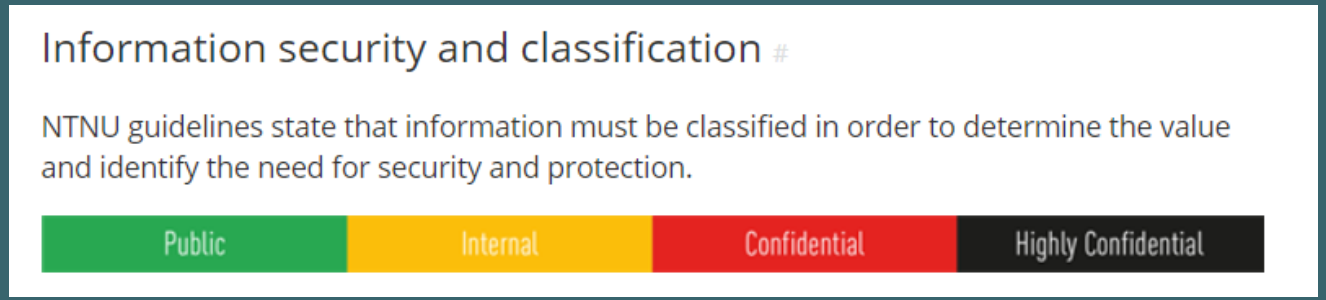


Special categories of personal data:

- Race or ethnic origin
- Political opinions, religion, philosofical beliefs, union memebership
- Sexual orientation and activity
- Health data
- Criminal offence data
- Genetic and biometric data (where used for identification purposes), examples include fingerprints, DNA, voice, gait etc.

Storage: Information Security

- All information (including research data) should be classified, to help choose the correct storage
  - Open, Internal, Confidential, Strictly confidential



https://innsida.ntnu.no/wiki/-/wiki/Norsk/Informasjonsklassifisering+-+informasjonssikkerhet

Storage guide

What physical storage media can I use?

Physical storage media refers to local storage and processing of information, for instance your computer (Mac, PC or hard drive).

Information classification:	Public	Internal	Confidential	Highly confidential
Privately owned laptop	**OK**	NO	NO	NO
Privately owned desktop	**OK**	NO	NO	NO
NTNU-acquired desktop (self-administered)	**OK**	**OK**	NO	NO
NTNU-acquired laptop (self-administered)	**OK**	**OK**	NO	NO
NTNU-administered desktop – encrypted	**OK**	**OK**	**OK**	NO
NTNU-administerd laptop – encrypted	**OK**	**OK**	**OK**	NO
USB drive/external hard drive	**OK**	**OK**	NO	NO
USB drive/external hard drive - encrypted	**OK**	**OK**	**OK**(1)	**OK**(2)

(1) The data must be stored in encrypted form on the storage media and the password kept in a separate location. Read more about how to encrypt files.

(2) The entire drive/disc must be encrypted with a strong password (read more on how to make passwords). The password must be kept in a separate location.

Storage services and collaboration platforms

Storage services and collaboration platforms refer to cloud services or servers at NTNU.
Click on the different solutions for more information.

Information classification:	Public	Internal	Confidential	Highly confidential
Personal cloud storage (dropbox, google drive ++)	**OK**	NO	NO	NO
NTNU Home directory («M:-drive»)	**OK**	**OK**	**OK**	**OK (1)**
NTNU Shared directory (T:-drive, group, project, etc.)	**OK**	**OK**	NO	NO
NTNU-administered Dropbox (contact Orakel)	**OK**	**OK**	NO	NO
NTNU-Box	**OK**	**OK**	NO	NO
Office 365 (SharePoint, Teams, Onedrive)	**OK**	**OK**	**OK**(1)	NO
NTNU NICE-1 - Storage solution with added security	**OK**	**OK**	**OK**	**OK (1)**
HUNT Cloud	**OK**	**OK**	**OK**	**OK (2)**
UiO TSD	**OK**	**OK**	**OK**	**OK**
NIRD (tidligere Norstore, driftes av Uninett Sigma2)	**OK**	**OK**	NO	NO

(1) Data must be encrypted. Read more on how to encrypt O365 files using AIP here or how to encrypt other files with 7-Zip

(2) Risk level is assesed on individual basis, see the HUNT information page for more information.

https://innsida.ntnu.no/wiki/-/wiki/English/Data+storage+guide

STORAGE

Storage guide

Storage services and collaboration platforms

Storage services and collaboration platforms refer to cloud services or servers at NTNU.
Click on the different solutions for more information.

Information classification:	Public	Internal	Confidential	Highly confidential
Personal cloud storage (dropbox, google drive ++)	**OK**	NO	NO	NO
NTNU Home directory («M:-drive»)	**OK**	**OK**	**OK**	**OK (1)**
NTNU Shared directory (T:-drive, group, project, etc.)	**OK**	**OK**	NO	NO
NTNU-administered Dropbox (contact Orakel)	**OK**	**OK**	NO	NO
NTNU-Box	**OK**	**OK**	NO	NO
Office 365 (SharePoint, Teams, Onedrive)	**OK**	**OK**	**OK**(1)	NO
NTNU NICE-1 - Storage solution with added security	**OK**	**OK**	**OK**	**OK (1)**
HUNT Cloud	**OK**	**OK**	**OK**	**OK (2)**
UiO TSD	**OK**	**OK**	**OK**	**OK**
NIRD (tidligere Norstore, driftes av Uninett Sigma2)	**OK**	**OK**	NO	NO

(1) Data must be encrypted. Read more on how to encrypt O365 files using AIP here or how to encrypt other files with 7-Zip

(2) Risk level is assesed on individual basis, see the HUNT information page for more information.

STORAGE

Åpen | Intern | Fortrolig | Strengt fortrolig

https://innsida.ntnu.no/wiki/-/wiki/Norsk/Informasjonsklassifisering+-+informasjonssikkerhet

Confidentiality

+

Hvilke fysiske lagringsmedier kan jeg bruke?

Med fysiske lagringsmedier mener vi behandling og lokal lagring av informasjon, feks lagring på egen maskin (Mac, PC eller harddisk). Les mer om de ulike lagringsmediene ved å klikke på dem.

Informasjonsklassifisering:	Åpen	Intern	Fortrolig	Strengt fortrolig
Privat-eid bærbar datamaskin	OK	NEI	NEI	NEI
Privat-eid hjemmemaskin	OK	NEI	NEI	NEI
NTNU-anskaffet hjemmemaskin (egenadministrert)	OK	OK	NEI	NEI
NTNU-anskaffet bærbar datamaskin (egenadministrert)	OK	OK	NEI	NEI
NTNU-administrert desktop – kryptert	OK	OK	OK	NEI
NTNU-administrert bærbar datamaskin – kryptert	OK	OK	OK	NEI
Minnepinne/ekstern harddisk	OK	OK	NEI	NEI
Minnepinne/ekstern harddisk – kryptert	OK	OK	OK(1)	OK(2)

(1) Dataene må ligge krypert på lagringsmediet og passordet oppbevart et annet sted. Les mer om kryptering.

(2) Hele disken må være kryptert med et godt passord (les mer om hvordan lage passord). Passordet må være oppbevart et annet sted.

Lagringstjenester og samhandlingsplattformer

Med lagringstjenester og samhandlingsplattformer mener vi lagring i skytjenester eller servere på NTNU. Les mer om de ulike tjenestene og plattformene ved å klikke på dem.

Informasjonsklassifisering:	Åpen	Intern	Fortrolig	Strengt fortrolig
Personlig skytjeneste (dropbox, google drive ++)	OK	NEI	NEI	NEI
NTNU Hjemmeområde («M:-disk»)	OK	OK	OK(1)	NEI
NTNU Fellesområde (T-enhet, gruppe, prosjekt, osv)	OK	OK	NEI	NEI
NTNU-administrert Dropbox (kontakt orakel)	OK	OK	NEI	NEI
NTNU-Box	OK	OK	NEI	NEI
Office 365 (SharePoint)	OK	OK	OK(1)	NEI
NTNU Onedrive (Office 365 og Sharepoint)	OK	OK	OK(1)	NEI
NTNU NICE-1 - Lagringsområde med økt sikkerhet	OK	OK	OK	NEI
HUNT Cloud	OK	OK	OK(2)	NEI
UiO TSD	OK	OK	OK	OK
NIRD (tidligere Norstore, driftes av Uninett)	OK	OK	NEI	NEI

(1) Dataene må ligge krypert på lagringsmediet og passordet oppbevart et annet sted. Les mer om kryptering og passord.

(2) Risikonivå vurderes individuelt, se nærmere informasjon på HUNTs informasjonssider

https://innsida.ntnu.no/wiki/-/wiki/Norsk/Lagringsguide

Accessibility

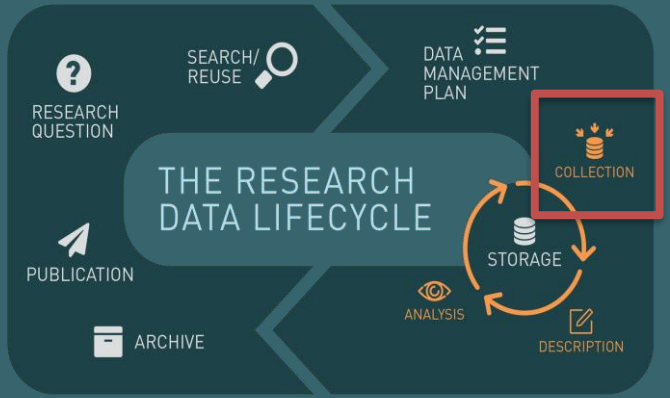
Integrity

Data collection

This guide is an overview of tools available at NTNU for collection of data, focused on collection of personal data, sound and video recordings. The overview will help you make correct choices for managing data in your research or student project.

Interview with recording of sound and/or video #

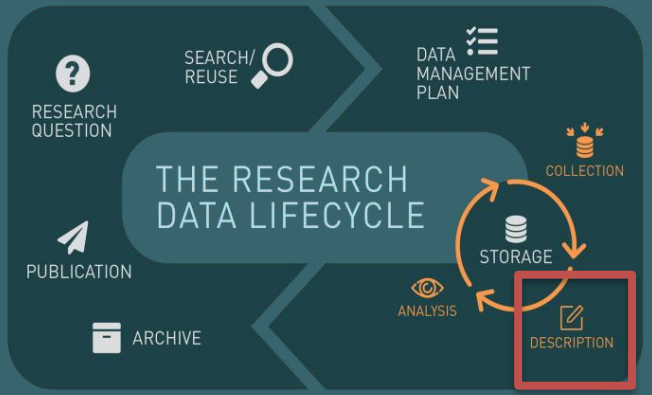
Information classification:	Public	Internal	Confidential	Highly Confidential
Zoom	**OK (1)**	**OK (1)**	NO	NO
Teams	**OK (1)**	**OK (1)**	NO	NO
Nettskjema-Diktafon App (X)	**OK**	**OK**	**OK (2)**	**OK (2)**
External dictaphone	**OK**	**OK**	**OK**	NO (3)



https://innsida.ntnu.no/wiki/-/wiki/English/Data+collection

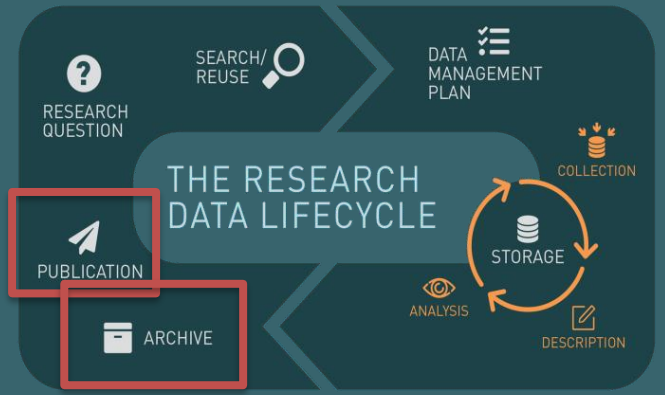
Describing data

- Make sure to keep all information that is necessary to understand and (re)use the data (both for others and your future self)
- **Metadata** and documentation should be developed during the project (much more work to do after the fact)
- Use standards where possible
- Name, structure and version files clearly and distinctly



Archive and/or publish data

- Archiving can be closed (restricted) or open (publishing)
  - Check requirements (funders, journal, institution)
  - Verifiability and reuse of data

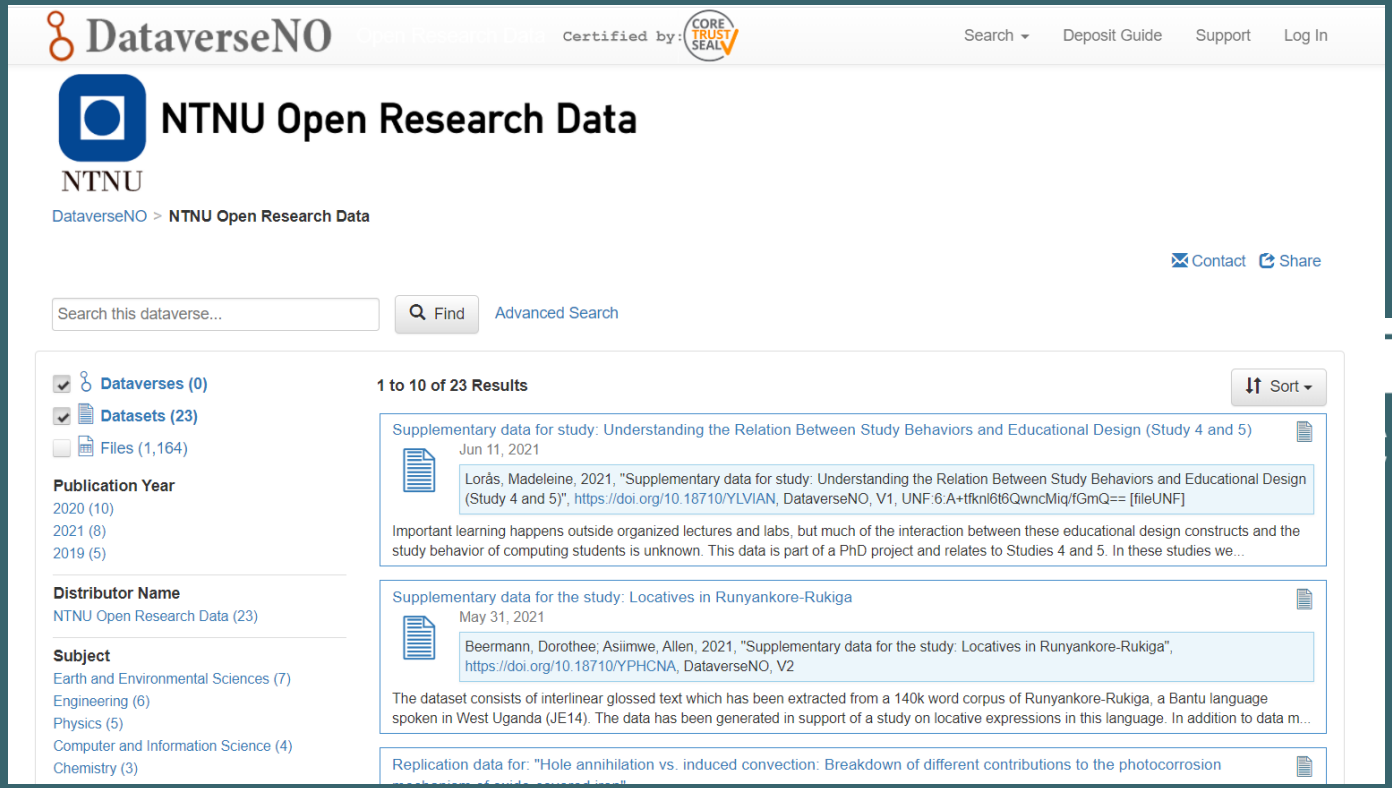


- More info and advice:

https://innsida.ntnu.no/wiki/-/wiki/Norsk/Arkivere+forskningsdata

Open data: How to publish?

- Data repository (recommended)
  - Subject specific
  - Institutional
    - NTNU Open Research Data (DataverseNO), BIRD
  - General
- Publishing in data journal (often in addition to repository)
  - Example: Scientific Data (Springer Nature)
- Supplement to paper or in publisher/journal repository
- Blog, ResearchGate, etc (preferably not…)



HIVING

https://dataverse.no/dataverse/ntnu

License information

Published under CC BY

This license lets others distribute, remix, tweak, and build upon your work, even for commercial purposes, as long as they credit you for the original creation. This is the most accommodating of licenses offered, and it is recommended for maximum dissemination and use of licensed materials.

Wind and Acceleration Data from the Hardanger Bridge

By **Aksel Fenerci**, **Knut Andreas Kvåle**, **Øyvind Wiig Petersen**, **Anders Rønnquist**, **Ole Øiseth**

https://doi.org/10.21400/5ng8980s

Published 18-08-2020 at Norges teknisk-naturvitenskapelige universitet 1587 views

The dataset consists of long-term wind and acceleration data collected from the Hardanger Bridge monitoring system. The data are collected through wind sensors (anemometers) and accelerometers that are installed on the bridge. The dataset includes both the raw data (in ".csv" format) and the organized data (with ".mat" extension based on hdf5 format). Downloadable zipped folders contain monthly data with different frequency resolutions, special events (storms, etc.) and the raw data. Details on the organization of the data can be found in the readme file and the data paper, both of which can be found in the dataset.

Resource type: Dataset

Category: Teknologi, Bygningsfag, Konstruksjonsteknologi

Process or method: GPS, Wi-Fi, accelerometers, anemometry, signal processing

Geographical coverage: Hardanger, Norway

acceleration | bridge | buffeting | complex terrain | dynamics | hardanger | monitoring | suspension | wind

ARCHIVING

https://doi.org/10.21400/5ng8980s

https://bird.unit.no/

NTNU

IPR: Intellectual Property Rights

- In general, NTNU retains rights to research results generated using NTNU's resources
- In commissioned/sponsored research, a formal contract regulates IPR
  - Academic freedom is still ensured
    - "NTNU's employees may not enter into agreements with third parties that violate the university's academic freedom and responsibility to make results from NTNU available so that they can be used as widely as possible in society and industry."

IPR and external partners

- Be aware of formal agreements/contracts
  - Confidentiality!
- Plan ahead (if possible)
  - Example: patents
- Resources and help:
  - https://innsida.ntnu.no/wiki/-/wiki/English/Intellectual+property+rights
  - https://innsida.ntnu.no/kommersialisering
  - https://www.ntnutto.no/patenting-and-ip/

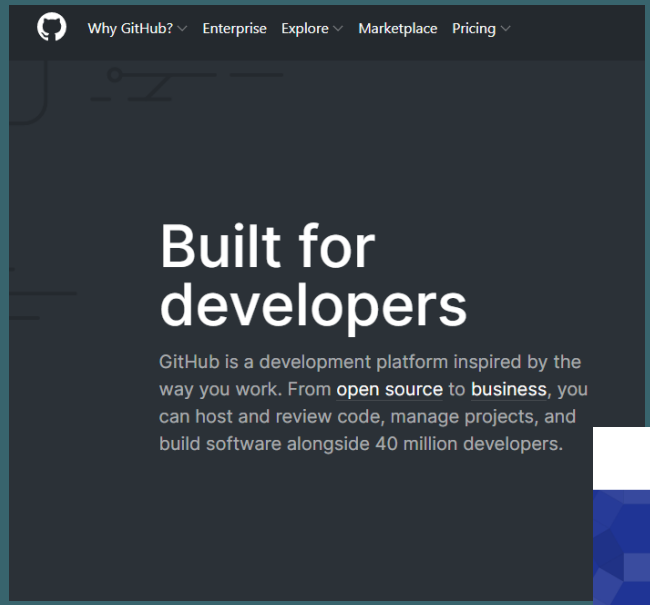
Open Source

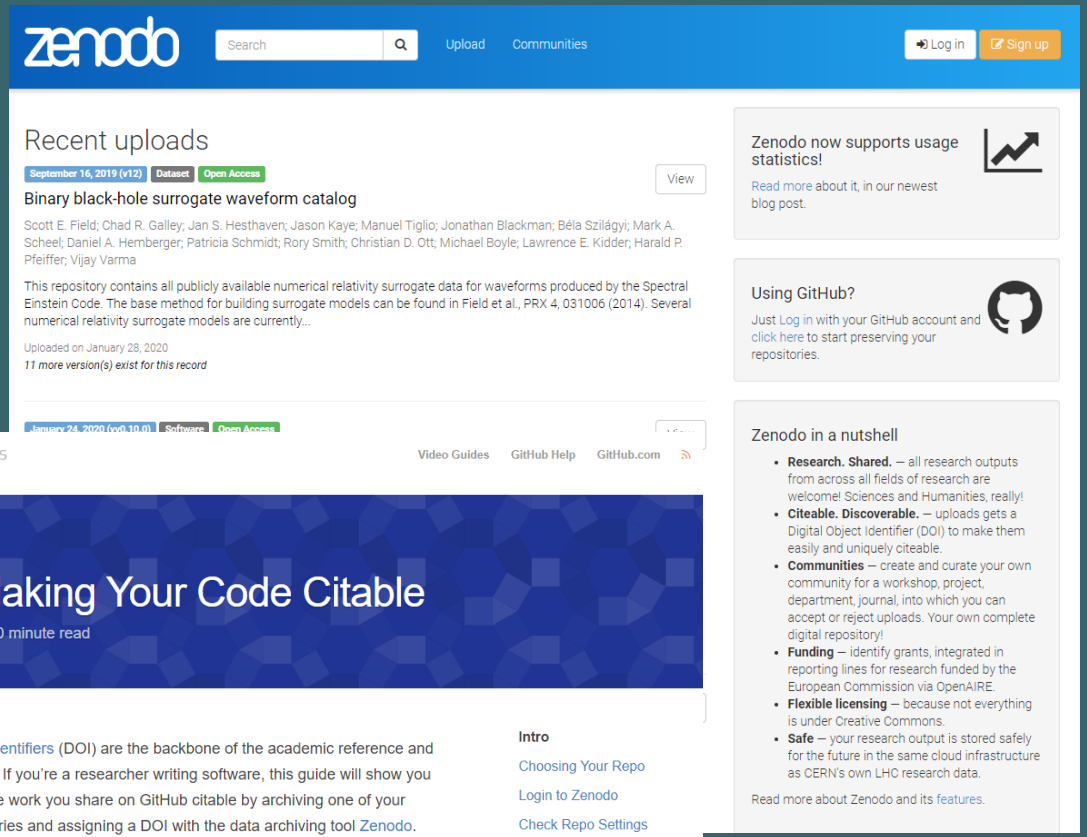
"Availability of the source code should soon become the minimum standard for academic software. In addition, culture should shift to embrace code review and appropriate credit for the developers of reusable software."

R. Bast, Nature Physics (2019) https://www.nature.com/articles/s41567-019-0624-3

- Code as part of published results
- Code review
- Software FAIRness
- Reproducible Research

- Also - Reproducible and open methods: https://www.protocols.io/


Why GitHub?
Enterprise
Explore
Marketplace
Pricing
Built for developers
GitHub is a development platform inspired by the way you work. From open source to business, you can host and review code, manage projects, and build software alongside 40 million developers.


zenodo
Search
Upload
Communities
Log in
Sign up
Recent uploads
September 16, 2019 (v12)
Dataset
Open Access
View
Binary black-hole surrogate waveform catalog
Scott E. Field; Chad R. Galley; Jan S. Hesthaven; Jason Kaye; Manuel Tiglio; Jonathan Blackman; Béla Szilágyi; Mark A. Scheel; Daniel A. Hemberger; Patricia Schmidt; Rory Smith; Christian D. Ott; Michael Boyle; Lawrence E. Kidder; Harald P. Pfeiffer; Vijay Varma
This repository contains all publicly available numerical relativity surrogate data for waveforms produced by the Spectral Einstein Code. The base method for building surrogate models can be found in Field et al., PRX 4, 031006 (2014). Several numerical relativity surrogate models are currently...
Uploaded on January 28, 2020
11 more version(s) exist for this record
Zenodo now supports usage statistics!
Read more about it, in our newest blog post.
Using GitHub?
Just Log in with your GitHub account and click here to start preserving your repositories.
Zenodo in a nutshell
Research. Shared. — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
Citeable. Discoverable. — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
Communities — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
Funding — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
Flexible licensing — because not everything is under Creative Commons.
Safe — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.
Read more about Zenodo and its features.

GitHub Guides
Video Guides
GitHub Help
GitHub.com
Making Your Code Citable
10 minute read
Digital Object Identifiers (DOI) are the backbone of the academic reference and metrics system. If you're a researcher writing software, this guide will show you how to make the work you share on GitHub citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool Zenodo.
ProTip: This tutorial is aimed at researchers who want to cite GitHub repositories in academic literature. Provided you've already set up a GitHub repository, this tutorial can be completed without installing any special software. If you haven't yet created a project on GitHub, start first by uploading your work to a repository.
Intro
Choosing Your Repo
Login to Zenodo
Check Repo Settings
Create a New Release
Minting a DOI
Finishing up

The Virtual Library - visit the library from your home office

The Virtual Library

Edit

Are you at home and in need of help with academic writing, citation, literature searches or loans? Visit The Virtual Library every weekday from 10am-2pm.

Norsk versjon: Det virtuelle biblioteket

THE VIRTUAL LIBRARY

NTNU | University Library

Table of Contents [-]

1. The waiting room
2. The main program
3. Course program
4. Contact

The NTNU University Library has a virtual library in Zoom for all students and staff at NTNU. Here you can talk to and get guidance from library staff, attend courses and get some real writing done at Shut up and write.

Visit The Virtual Library

NTNU IT – Research Support

RSE

Research Software Engineering

LAB

Laboratory Instrumentation and Support

Mime

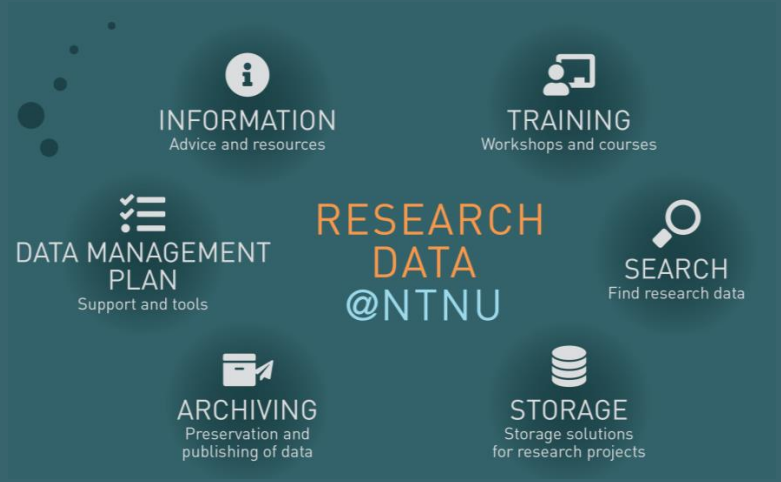
IT support for PhD students

HPC

High Performance Computing

Resources and support

- NTNU's pages about publishing: https://innsida.ntnu.no/publisering
- NTNU's pages about research data: https://innsida.ntnu.no/researchdata



- Questions (or feedback)?
  - Help from the library, in particular related to the publishing fund: publishing@ub.ntnu.no
  - Help with research data, from the library and IT: research-data@ntnu.no

Thank you for you attention!