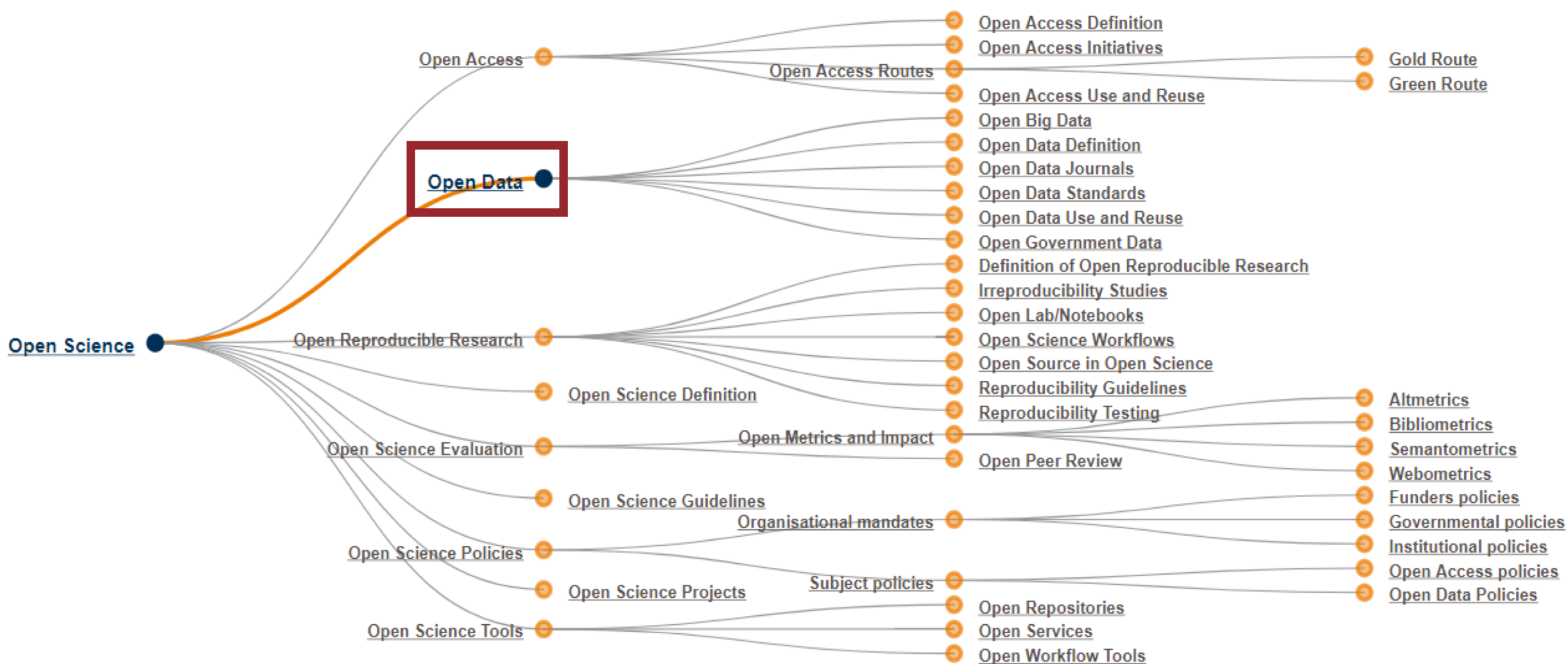# Open data – requirements from funding bodies

Presentation for ELEXIR

February 21th 2020

dr. lars figenschou

University Library, UiT The Arctic University of Norway

The context - Open Science....

Open Data are online, free of cost, accessible data that can be used, reused and distributed provided that the data source is attributed. (Foster Open Science)

https://www.fosteropenscience.eu/foster-taxonomy/open-data

The context - Open Data....

"We have an obligation to save data for the future!"

Kenneth Ruud, Munin Conference 2017

PRINCIPLES AND GUIDELINES
for research data management at UiT (from sep. 2017)

Purpose:
…to clarify responsibility and provide guidance about how the institution and its employees shall manage, share, and archive research data in line with the institution's administrative, financial, and ethical guidelines.

Should not be an(other) administrative burden, but part of good research practice!!

PRINCIPLES AND GUIDELINES
for research data management at UiT

Research data is…:
all registrations, notes, and reporting which are produced or arise in the course of research, and which are regarded as being of scientific interest and/or scientific potential. The format of these may include, but is not limited to, numbers, text, source code, photographs, films, and sound.

National strategy…:
Institutions must decide which data is worth archiving and make those reusable.

PRINCIPLES AND GUIDELINES
for research data management at UiT

As a general rule, UiT owns all research data produced by employees at UiT.

The researcher shall ensure that the research data is securely stored, backed up, and archived, in accordance with the Information security management system at UiT

The researcher shall write a data management plan in an early phase of the project and preferably within six months of the start of the project.

What is a DMP?

- A Data Management Plan describes how research data will be collected, processed, stored and made accessible, and preferably also how much research data management will cost.

- The DMP is being created at the start of the project, but it is revised along the way.

- NB: DMP requirements apply to all people at UiT who handle research data

DMP: Content

- Project Information

- Responsibilities and rights

- Collection/generating data

- Documentation and metadata

- Storage and preservation during the project period

- Archiving and sharing

- Ethics and consent

Why is DMP important?

Morgan Edwards
@mangoedwards

Følg

I can't send you the original data because I don't remember what my excel file names mean anymore #overlyhonestmethods

RETWEETS
129

LIKER
80

09.11 - 8. jan. 2013

Why is DMP important?

Retraction Watch

Tracking retractions

Doing the right thing: Authors retract brain paper with "systematic human error in coding"

with one comment

A group of Swiss neurologists have lost their 2013 article in *Frontiers in Human Neuroscience* after reporting that their data were rendered null by coding errors.

frontiers in
HUMAN NEUROSCIENCE

2.9

Research data @ UiT

PhD regulations at UiT

(remember what the UiT policy says, and complete with):

"The student shall set up a Data Management Plan for research data in the project within six month according to the current regulations."

"The committee may request that the candidate submit his/her data and any additional information that complements or clarifies the thesis."

From: *Regulations concerning the degree of Philosophiae Doctor (PhD) at the University of Tromsø - The Arctic University of Norway (UiT).*

PRINCIPLES AND GUIDELINES
for research data management at UiT

As a general rule, UiT owns all research data produced by employees at UiT

The researcher shall ensure that the research data is securely stored, backed up, and archived, in accordance with the Information security management system at UiT

The researcher shall write a data management plan in an early phase of the project and preferably within six months of the start of the project
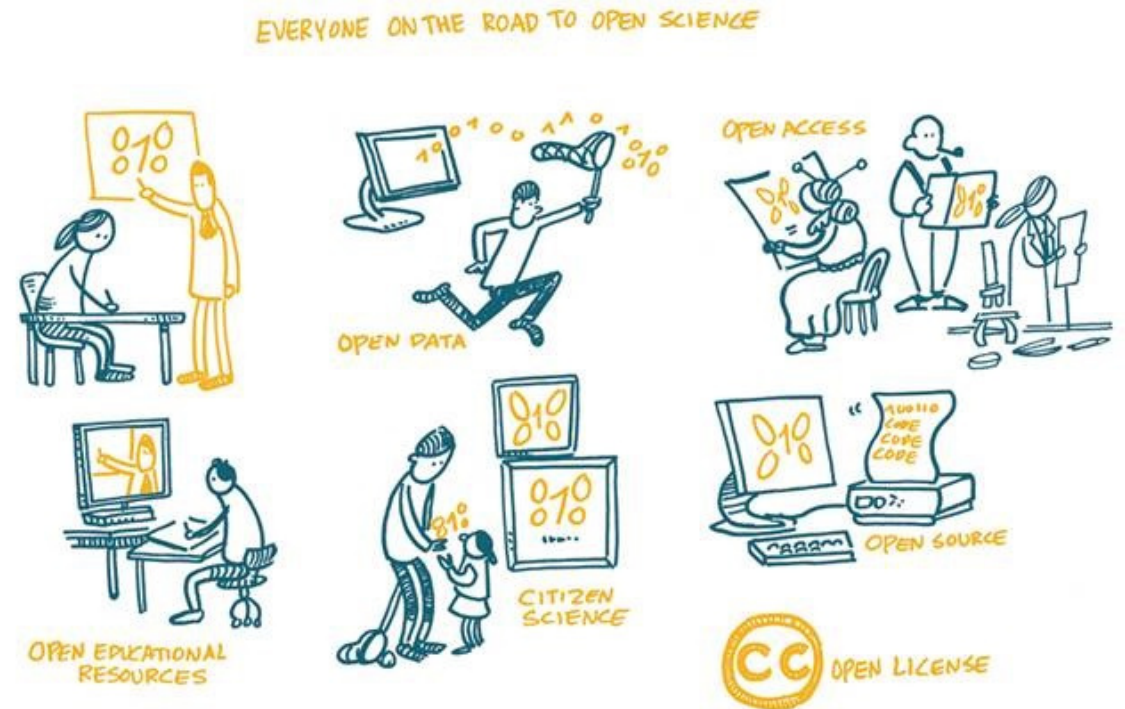
The researcher shall make the research data openly available for future use by all relevant users, providing this is not prevented by any legal, ethical, security, or commercial reasons.

Research data shall be equipped with standardised metadata that enable other researchers to search for - and use the data.

Open data – best practices…

- Research data should be accessible and reusable.

- They therefore need to be findable via as many routes as possible.

- To be re-usable the data must be understandable.

- To read and interpret the data you must be able to open the files.

The FAIR Data Principles

- **F**indable
  The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.

- **A**ccessible
  Once the user finds the required data, she/he needs to know how they can be accessed, possibly including authentication and authorization.

- **I**nteroperable
  The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- **R**eusable
  The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

Benefits for researchers and ELEXIR

Researcher:

- Enhance the visibility of your research – more connections and collaboration opportunities.
- Make the data citable by minting DOI to the data – more citations.
- Long term preservation of the data.
- Increase credibility and recognition through transparency.
- Ensure compliance with funding agency mandates and journal publishing policies.

ELEXIR:

- Fulfill the obligation to save data for the future.
- A separate and citable collection for archived and published data.
- Increased visibility and awareness.
- The legacy of ELEXIR will continue to be visible - in the long term.

Sum it up... Expectations and requirements

UiT The Arctic University of Norway

UiT will be in the national forefront in Open Science and our research data and publications will be openly available when possible.

- All projects that involve research data must have a data management plan.

- Research data shall be stored and archived at UiT-approved facilities/repositories.

- Research data shall be made openly accessible for further use provided that there are no legal, ethical, security or commercial reasons for not doing so.

- The research data shall be made openly accessible as early as possible. For data that form the basis of scientific publications, this means no later than the date of publication.

- Research data shall be equipped with standardized metadata and licenses that enable other researchers to search for and use the data.

[Principles and guidelines for research data management at UiT](#)

Expectations and requirements (NFR/RCN & Ministry of Education and Research)

The Research Council of Norway

Policy on open access to research data:

- The guidelines apply to all data generated by projects funded by the Research Council – with a few exceptions.

- All projects awarded funding must assess the need for a data management plan.

- The research data that are stored must be of quality that makes them possible to find and reuse.
  The Research Council recommends that you follow the international FAIR principles.

Ministry of Education and Research

National Strategy on access to and sharing of research data (2017).

Fundamental principles:

- Research data must be as open as possible and as closed as necessary.

- Research data should be managed in a way that take care of the values in the data and make them available for utilization to the best possible extent.

Expectations and requirements (EU)

- Horizon 2020:
  - «As open as possible, as closed as necessary».
  - From 2017 – Data Management Plans are mandatory

- Horizon Europe:
  - more reliable science (by allowing data verification)
  - more efficient science (by sharing resources)
  - foster research integrity
  - introduce openness as criterion for receiving research funding

Expectations and requirements

- Mandate research data management in line with FAIR data principles, and promote long-term preservation.

- Improve the practice of research and innovation:
  - Early sharing of all research outputs
  - All data FAIR
  - Reproducible results
  - A research system that rewards and incentivizes researchers to adopt Open Science practices, with appropriate metrics.

Expectations and requirements – Journals

After publication, all data and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of a *Science* Journal. After publication, all reasonable requests for data, code, or materials must be fulfilled. Any restrictions on the availability of data, code, or materials, including fees and restrictions on original data obtained from other sources must be disclosed to the editors as must any Material Transfer Agreements (MTAs) pertaining to data or materials used or produced in this research, that place constraints on providing these data, code, or materials. Patents (whether applications or awards to the authors or home institutions) related to the work should also be declared.

Fossils or other rare specimens must be deposited in a public museum or repository and available for research.

Unreasonable restrictions on data, code, or material availability may preclude publication.

can be reported at **science_data@aaas.org**.

Expectations and requirements – Journals

- **Journal of the Royal Society –** To allow others to verify and build on the work published in Royal Society journals, it is a condition of publication that authors make available the data, code and research materials supporting the results in the article.
Datasets and code should be deposited in an appropriate, recognized, publicly available repository.

- **Nature –** Authors are required to make materials, data and associated protocols promptly available to others without undue qualifications. Data sets must be made freely available to readers from the date of publication, and must be provided to editors and peer-reviewers at submission, for the purposes of evaluating the manuscript.

Open data – prepare for the future!!

- New project or application?
  - A plan for research data management will ease the burden, raise the quality and increase the efficiency of the work.
  - Make data deposit an integrated part of the project.
  - Sharing by archiving and publishing.
  - Make your data citable and cite your data.
- Old data
  - Start with the easiest part – publish background data for a publication.
  - Data without description has little or no value.

Research Data Management and Open Data

- ✓ *A key factor for success is to make a data management plan.*

- ✓ *Make data deposit an integrated part of the project.*

- ✓ *Deposit data as early as possible – makes it easier for the researchers.*

- ✓ *Follow international standards – FAIR Data Principles!*

Kenneth Ruud at the 12th Munin Conference on Scholarly Publishing 2017

- ✓ *Assigning DOI to data makes data* citeable *– data gives credit to the researchers.*

- ✓ *We have an obligation to save data for the future!*

- ✓ *With DORA, there will be a broader look on scientific impact – including research data.*

- ✓ *Research data might be the most important publication in the future.*

PART II

Dataverse.no
https://dataverse.no

Developed at Harvard - OPEN SOURCE

UiT Open Research Data
https://dataverse.no/dataverse/uit

54 Installations

Open data @ UiT

Three main types of research data repositories

- Domain-specific repositories, e.g. BioModels, Database of Genomic Variants Archive

- Institutional repositories, e.g. UiT Open Research Data

- General purpose repositories, e.g. Zenodo, Figshare, Harvard Dataverse

Within each category, there a many repositories available, and researchers have to relate to different repositories in the course of their career.

Most important recommendation: Choose a **trusted** repository, that is one that is aligned with the **FAIR principles** of data management and sharing, and most importantly, one that makes your data **discoverable**.

How to make your data FAIR?

How does archiving and publishing your data in a trusted data repository make your data FAIR?

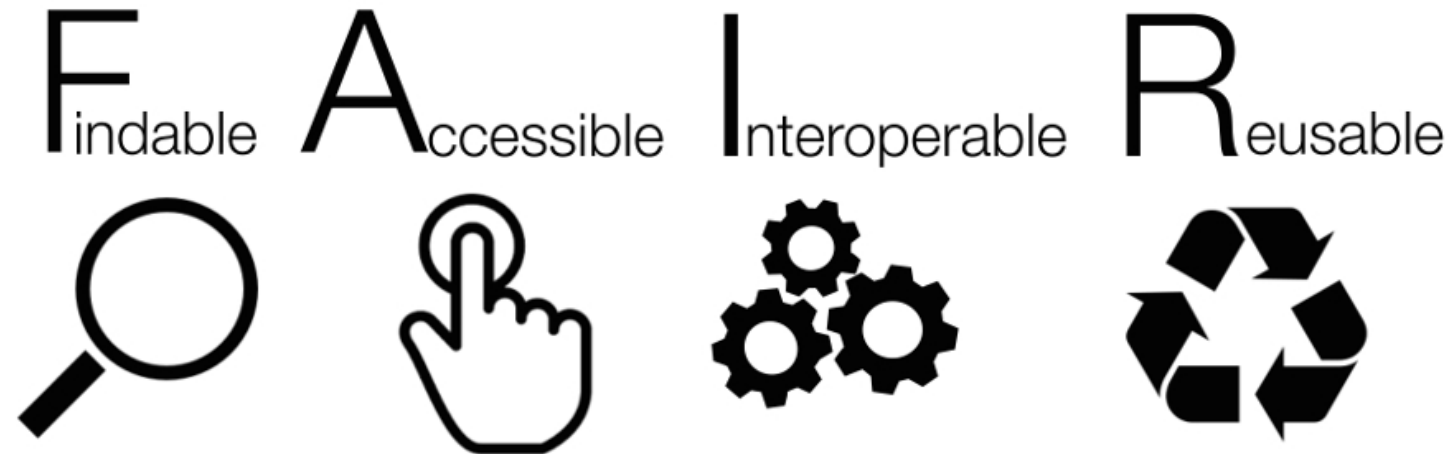Findable   Accessible   Interoperable   Reusable

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

Findable

- Published with a persistent identifier
- Good metadata
- Indexed in search engines

Findable

Metadata = description of data
Examples of metadata:
- Author, **title**, **description**, …
- **Keywords**
- Geographical information

Keyword ❓   Methane
Dissolution
Model
Gas
Bubbles

Geospatial Metadata ⌃

Geographic Coverage ❓   Svalbard and Jan Mayen

Geographic Bounding Box ❓   westlimit=9.333333 eastlimit=9.6(
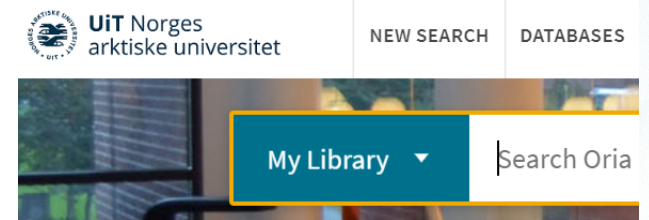
Jansson, Pär; Ferre, Benedicte; Silyakova, Anna; Dølven, Knut Ola; Omstedt, Anders, 2019, "Replication Data for: A new numerical model for understanding free and dissolved gas progression towards the atmosphere in aquatic methane seepage systems", https://doi.org/10.18710/LS2KUX, DataverseNO, V2

DOI = Digital Object Identifier = a type of persistent identifier ~ persistent URL

DataCite
FIND, ACCESS, AND REUSE DATA

Google Dataset Search Beta

BASE
Bielefeld Academic Search Engine

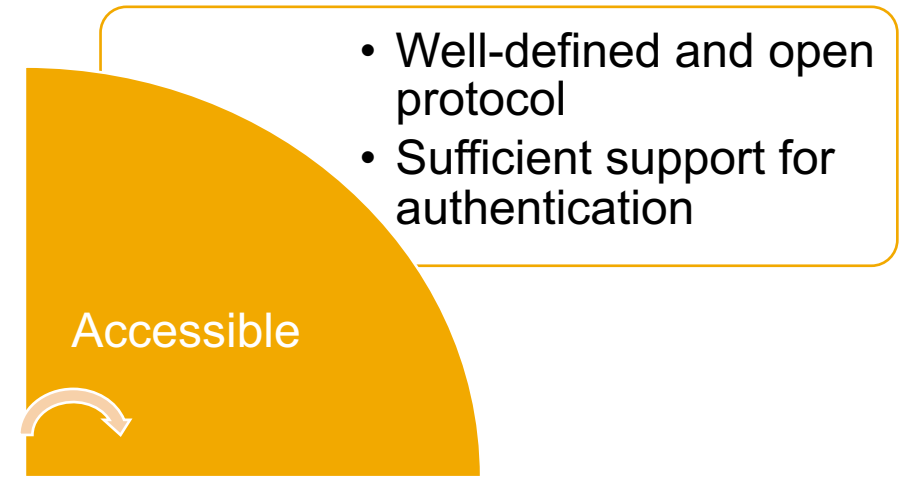UiT Norges arktiske universitet   NEW SEARCH   DATABASES

My Library ▼   Search Oria

Accessible

- Published data should be accessible through a well-defined and open protocol (e.g. https).

- If necessary, the protocol must enable sufficient authentication of users, e.g. for access to sensitive data.

Accessible

- Well-defined and open protocol
- Sufficient support for authentication

Interoperable

- Metadata are based on common metadata standards This applies to both
  - general metadata, e.g. international date format (e.g. ISO-8601): YYYY-MM-DD (2019-12-09), and
  - domain-specific metadata, e.g. Darwin Core = a standard for description of data on biological diversity.

- Keywords are based on controlled metadata vocabularies, e.g. Darwin Core standard values for age class or life stage of a biological individual:

- Interoperability enable search and re-user across datasets and repositories.

lifeStage	
Identifier	http://rs.tdwg.org/dwc/terms/lifeStage
Definition	The age class or life stage of the biological individual(s) at the time recorded.
Comments	Recommended best practice is to use a controlled vocabulary.
Examples	egg , eft , juvenile , adult

13.1.2020

(cc) BY

Re-usable

- Data are documented, so that your peers understand them and are able to (re-)use them; e.g. in a ReadMe file.

- Data are archived in preferred/sustainable file formats, so that the files can be opened and read also in the long term; e.g. tabulator-separated plain text (.txt) for Excel spreadsheets (.xlsx).

- Data are furnished with a clear use license, so that researchers who want to use them know what they are allowed to do with them.
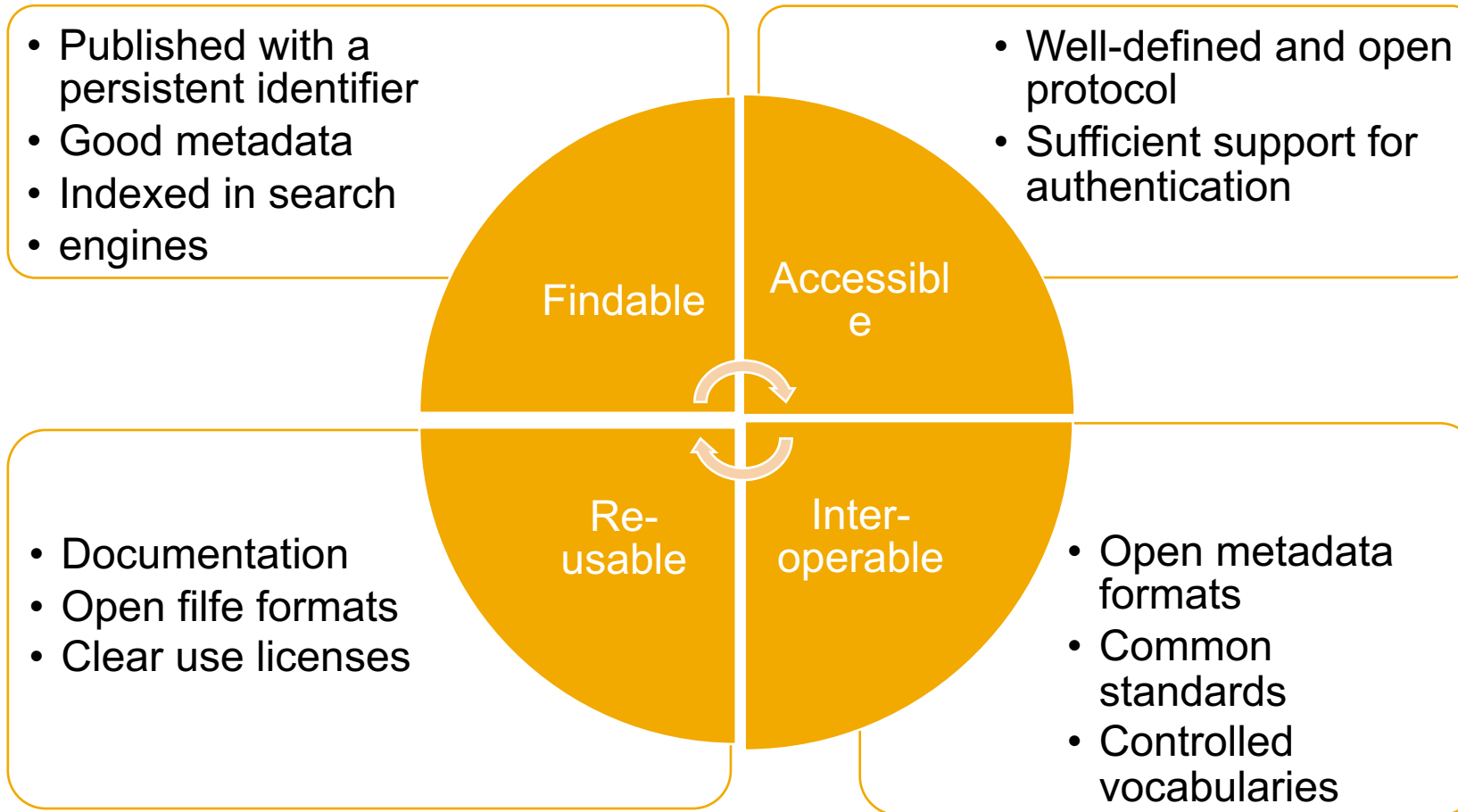
Example: Creative Commons (CC) licenses

- Documentation
- Open filfe formats
- Clear use licenses

Re-usable

Together: **FAIR**

- Published with a persistent identifier
- Good metadata
- Indexed in search
- engines

- Well-defined and open protocol
- Sufficient support for authentication

Findable

Accessible

Re-usable

Inter-operable

- Documentation
- Open filfe formats
- Clear use licenses

- Open metadata formats
- Common standards
- Controlled vocabularies

Dataverse.no or UiT Open Research Data……
Limitations?

- As of today, UiT Open Research Data may not be suitable for archiving **all raw** data.

- We are working on improving support for larger amounts of data.

- But, bear in mind: The problem is often not **storage**, but **re-use**, that is how others may access (download?) and analyse your data.

- **Selection of data** for sharing: Which data and how much data are needed to carry out research?; e.g. a selection of raw data?; only processed data? → This is for the research community to decide (cf. *National Strategy on access to and sharing of research data*, Ministry of Education and Research)

Data are first-class research outputs

General recommendation:

- Treat your data as first-class research outputs – on the same level as your other types of publication, e.g. articles.

- Data are research outputs in their own right, not just attachments to articles.

- Data should have their own reference for citation.

Dataset

Today

…many consider the dataset as (some kind of) supplement…..

Future

…Research data might be the most important publication in the future.
that is, ***open data will be the important unit – not the article***

Support and Training

UiT provides research data management support throughout the lifecycle of your project(s):

- General information about FAIR research data management is collected on the UiT Research Data Portal (https://uit.no/researchdata).

- Training session on different topics

- Feedback on data management plans

- Institutional repository: UiT Open Research Data (http://opendata.uit.no/)

- Advice on how to establish and manage larger collections

- Curation of datasets to be published (= quality check of FAIR alignment)

(cc) BY

More info about research data management

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

E-mail Support: research-data@support.uit.no

UiT Open Research Data: http://opendata.uit.no/