

Practical information

- 15 minutes breaks between sessions
- Remember to mute when listening
- Use chat actively during sessions
- Use “raise hand” functionality during discussions or to indicate wish to ask questions
- No recording of sessions are planned
- Working groups are organised using the breakout session functionality in Zoom
 - Participants are allocated to sessions by host
- Supposed to be an interactive course!
- Picture of all participants wanted, will start with this

An illustration of an iceberg floating in a blue ocean. The tip of the iceberg is above the water surface, while the much larger, jagged base is submerged. The text is overlaid on the iceberg.

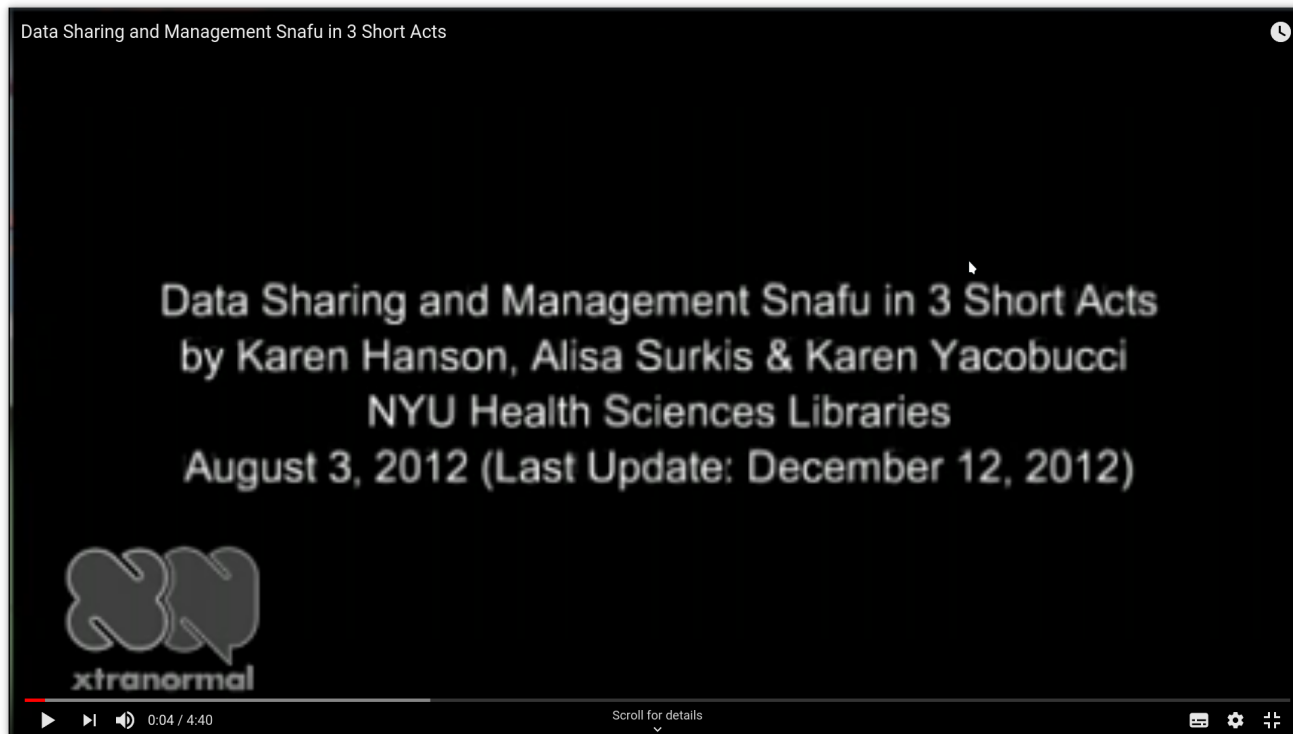
Motivation: Why do we need data management?

Øystein Godøy

Outline

- Data Sharing and Management Snafu in 3 Short Acts
 - <https://www.youtube.com/watch?v=N2zK3sAtr-4>
- Why do we need data management?
- Science life cycle/Data life cycle
- How to change data sharing culture.
- What are the FAIR data principles?
 - How do they help with good data management?
- External boundary conditions by funding agencies and publishers.
- Scientific data as service.
- Data management plan.

Data Sharing and Management Snafu in 3 Short Acts



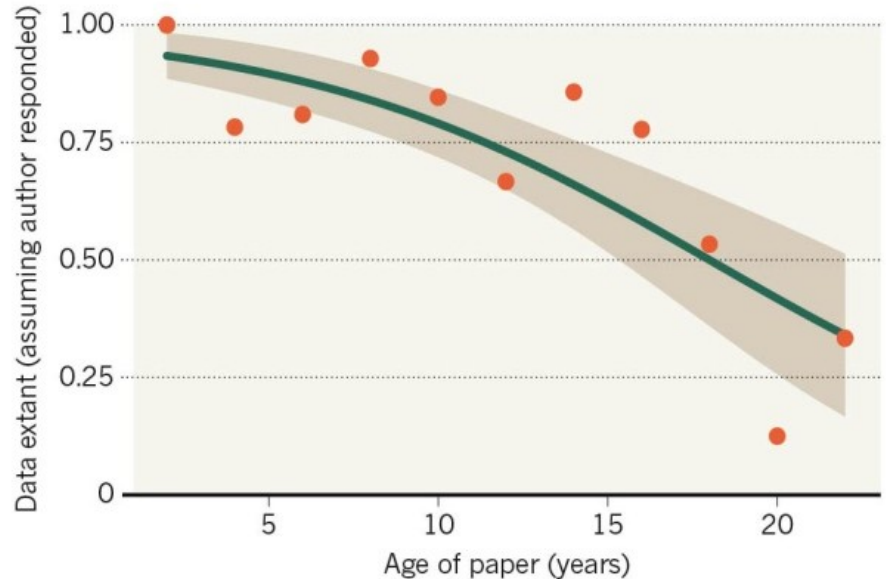
<https://www.youtube.com/watch?v=N2zK3sAtr-4>

Why do we need data management?

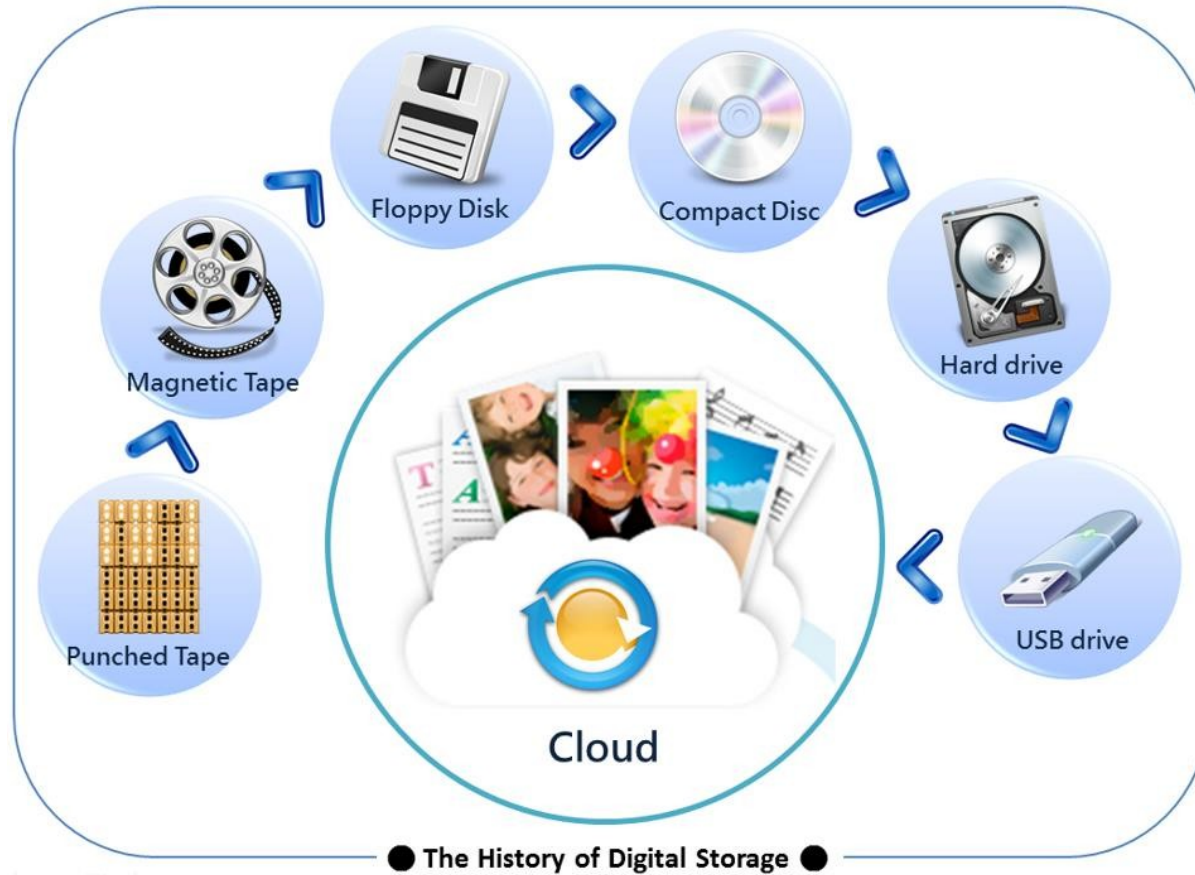
- Loosing scientific data
 - Decline can mean 80% of data are unavailable after 20 years.
 - Gibney and Van Noorden (2013), Nature

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.







Why bother with structured data management?



- Benefits
 - Maximise public investment in data collection and production
 - Promote scientific collaboration
 - Promote interdisciplinary science
 - Promote scientific transparency
 - Leave a legacy
- Science paradigms
 - according to Jim Gray
 - empirical science
 - theoretical science
 - computational science
 - data exploration science

Why share data?

- Research sponsor require it
 - recognition as an authoritative source and wise investment
- Quality control
 - improved data quality due to expanded use, field checks, and feedback
- Improved visibility
 - improved connections to scientific network, peers, and potential collaborators
- Journals require it
 - Reproducible research
- Far upstream sponsors require it



CC image by SLU Madrid Campus
on Flickr

Making Your Research Easier and Cheaper

The 5 P's matter!

Prior

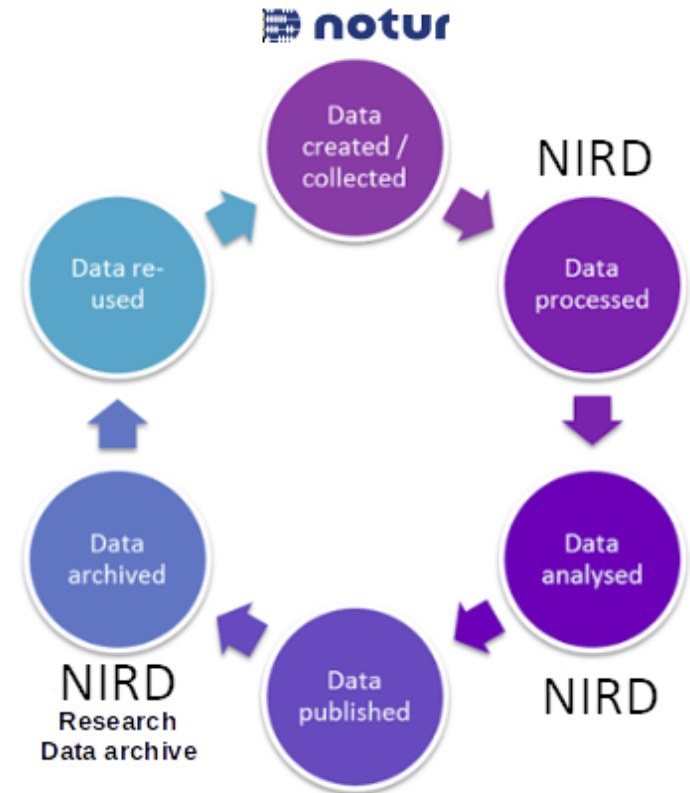
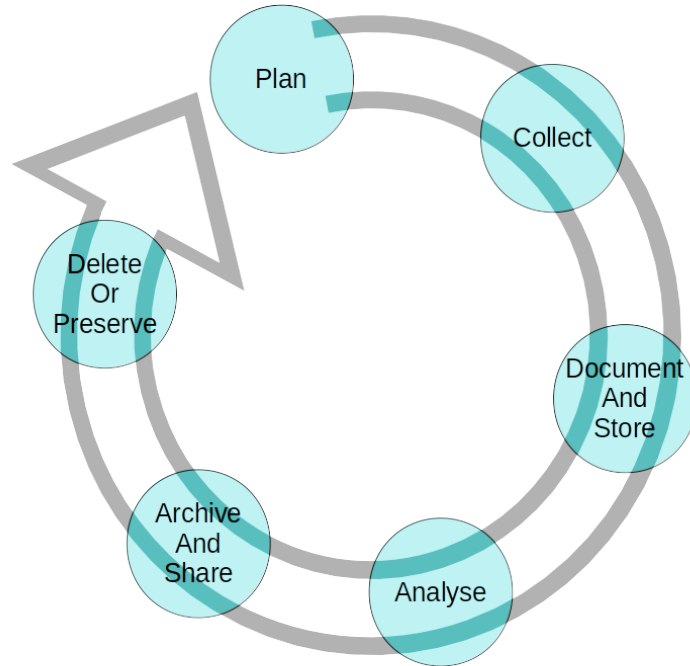
Planning

Prevents

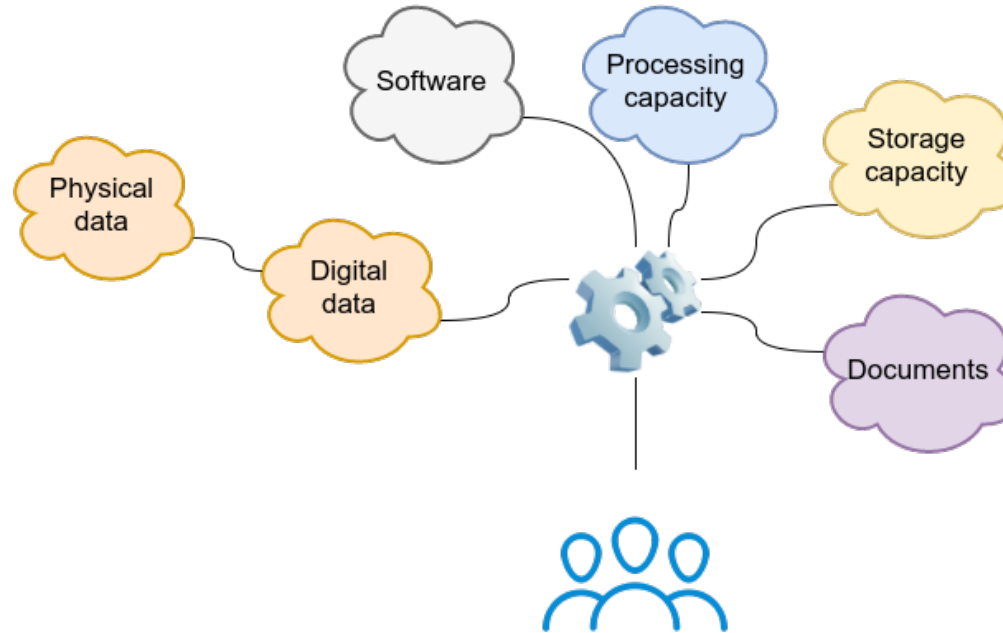
Poor

Performance!

Science life cycle/Data life cycle

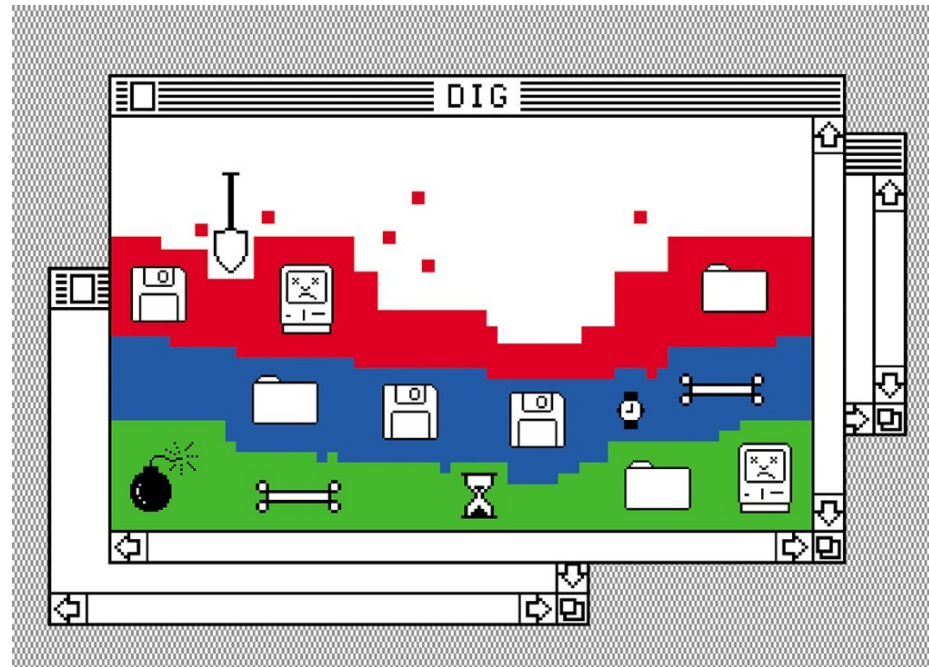


<https://codemeta.github.io/>



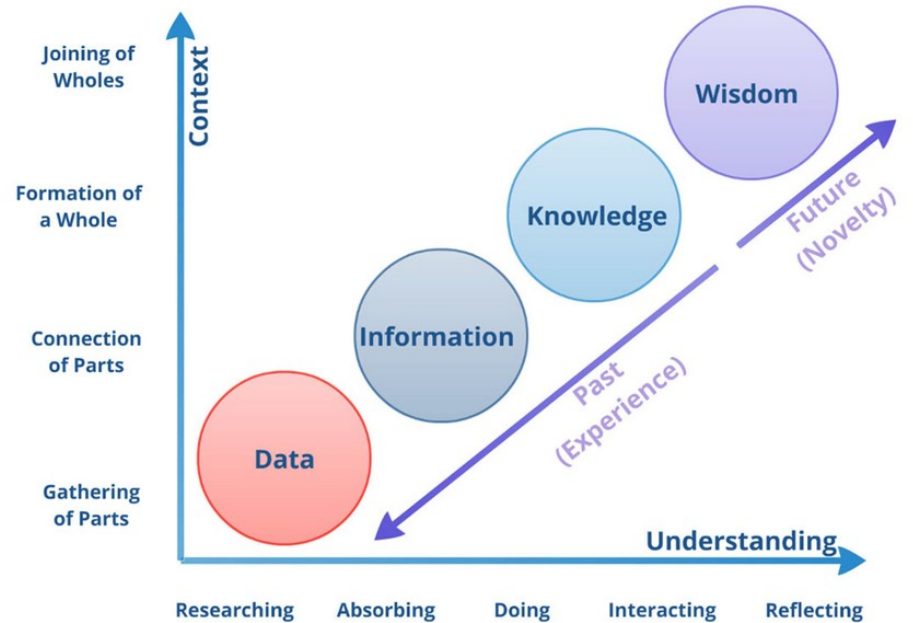
Challenge to scientists: does your ten-year-old code still run?

<https://www.nature.com/articles/d41586-020-02462-7>



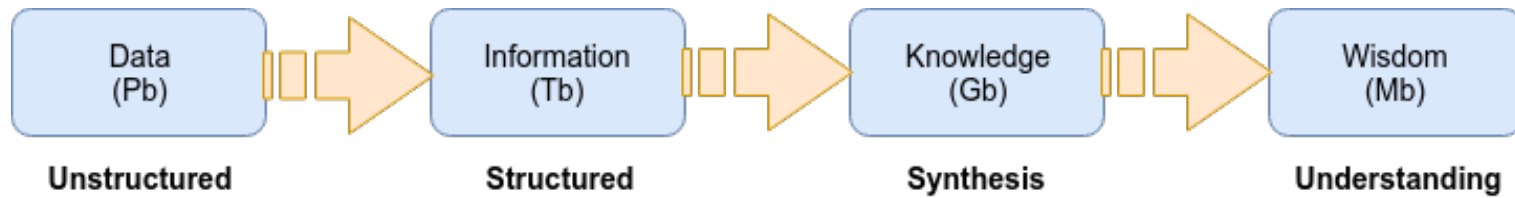
DIKW chain

- How to transition from data to knowledge and understanding...
 - The illustration is a common redrawing of Russ Ackoff “From Data to Wisdom”
 - Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9
- DIKW is necessary to
 - Take care of data for the future
 - Ensure data is the basis for knowledge
 - Now and in the future
 - Knowledge based management depends on national, regional and global interaction



<http://www.easterbrook.ca/steve/2012/09/what-is-climate-informatics/>

DIKW chain



The reality today

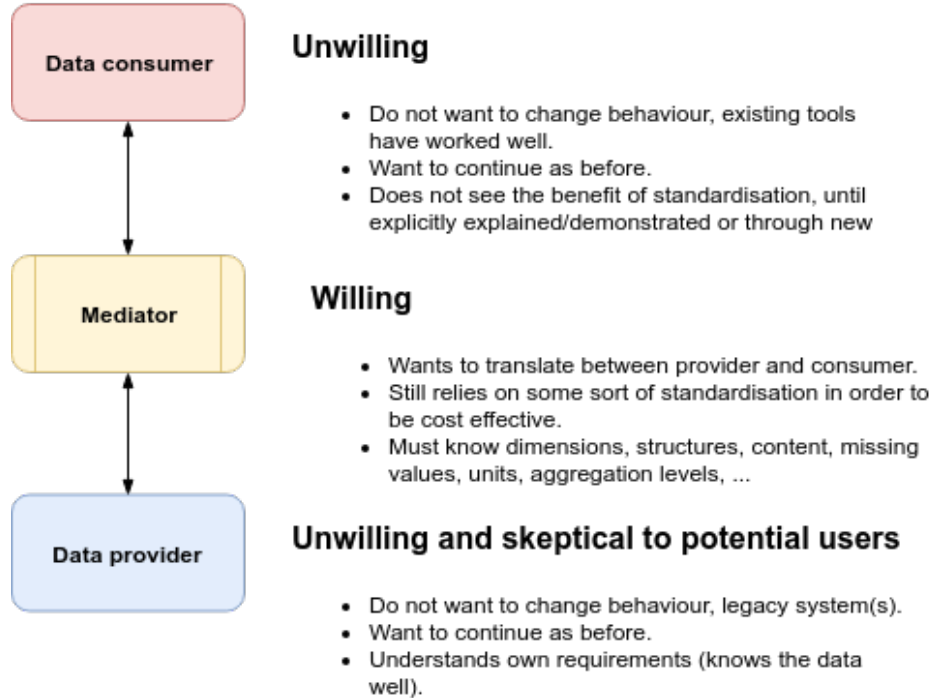


Recreated from Klump et al. 2006

How to change data sharing culture?



Bridging between data management actors



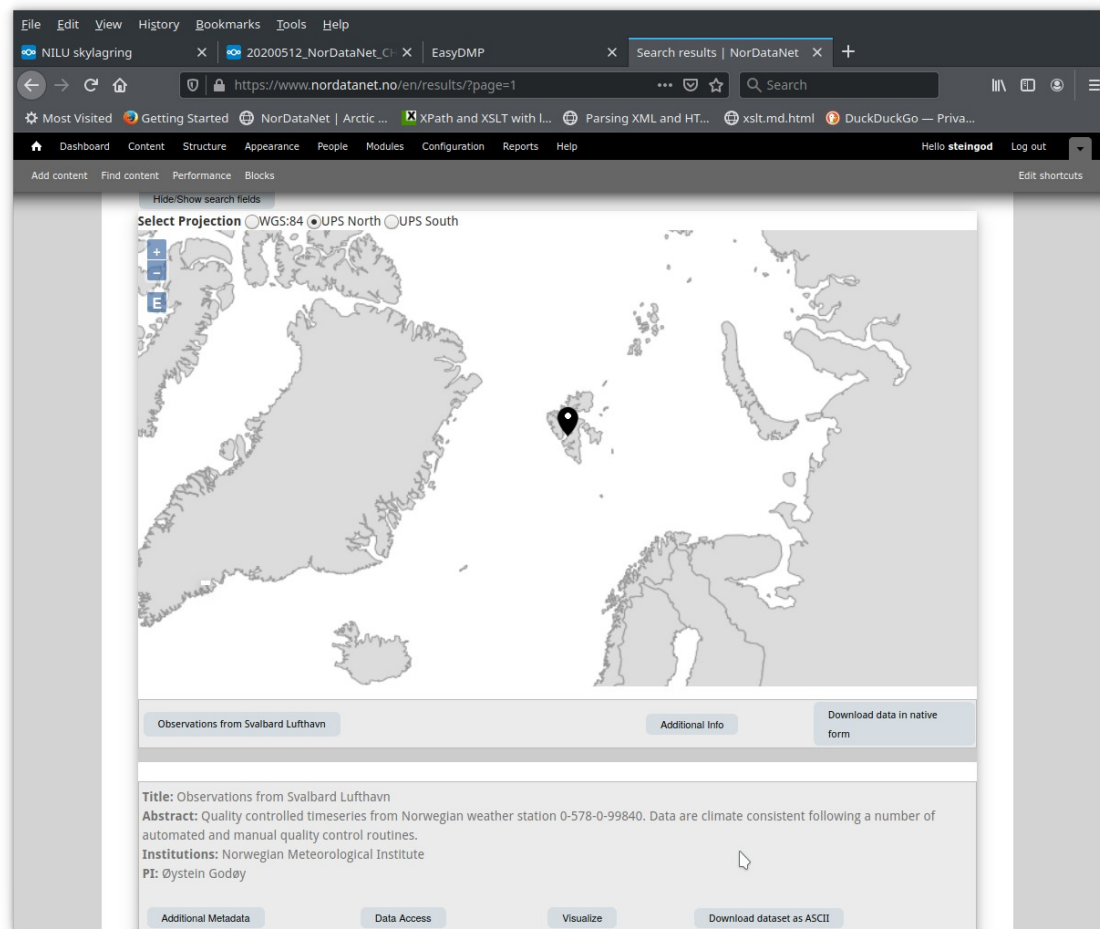
- A prerequisite for efficient data sharing across communities is application of proper metadata and standards
- Although standards exist, they are often not used by data providers who doesn't see the benefit
- Lacking understanding for the importance of use metadata
 - Enabling reuse across communities and generations
 - Lacking understanding for the importance of semantic standardisation
- Need a business model crediting all involved parties
 - Scientists, institutions, data centres,
- It is about leaving a legacy

The FAIR Guiding Principles for scientific data management and stewardship

- To be **Findable**:
 - F1. (meta)data are assigned a *globally unique and persistent identifier*
 - F2. data are described with *rich metadata* (defined by R1 below)
 - F3. metadata clearly and explicitly include the identifier of the data it describes
 - F4. (meta)data are *registered or indexed in a searchable resource*
- To be **Accessible**:
 - A1. (meta)data are retrievable by their identifier using a *standardized communications protocol*
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
 - A2. metadata are accessible, even when the data are no longer available
- To be **Interoperable**:
 - I1. (meta)data use a *formal, accessible, shared, and broadly applicable language for knowledge representation*
 - I2. (meta)data use *vocabularies* that follow FAIR principles
 - I3. (meta)data *include qualified references* to other (meta)data
- To be **Reusable**:
 - R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible *data usage license*
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data *meet domain-relevant community standards*

Scientific data as service

- From science to service
 - Goddard, Science 23 Sep 2016: Vol. 353, Issue 6306, pp. 1366-1367 DOI: 10.1126/science.aag3087



The screenshot shows a web browser window displaying the NorDataNet search results page. The browser's address bar shows the URL <https://www.nordatanet.no/en/results?page=1>. The page features a map of the Arctic region with a location pin on Svalbard. Below the map, there are buttons for 'Observations from Svalbard Lufthavn', 'Additional Info', and 'Download data in native form'. The main content area displays the following information:

Title: Observations from Svalbard Lufthavn
Abstract: Quality controlled timeseries from Norwegian weather station 0-578-0-99840. Data are climate consistent following a number of automated and manual quality control routines.
Institutions: Norwegian Meteorological Institute
PI: Øystein Godøy

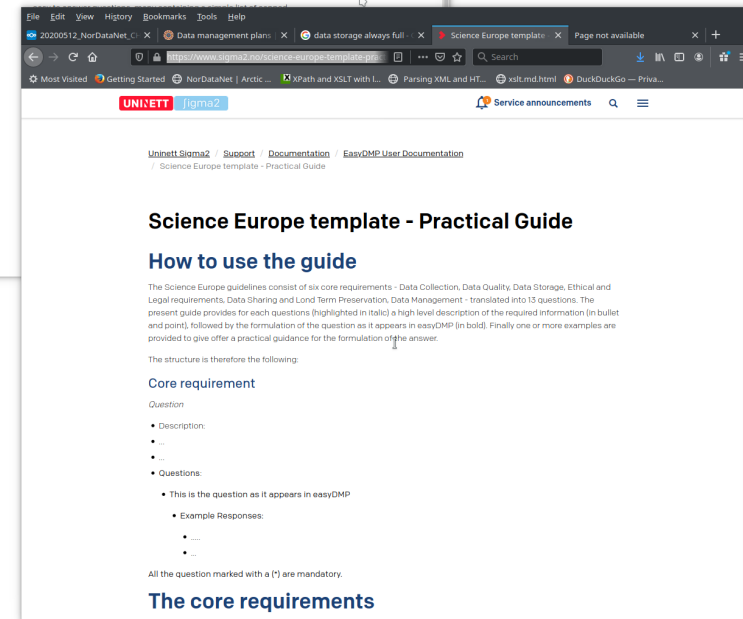
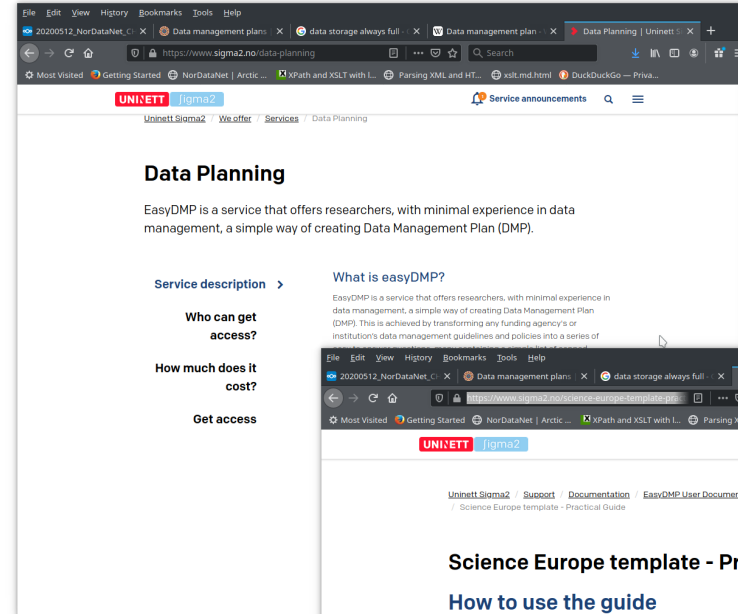
At the bottom of the page, there are buttons for 'Additional Metadata', 'Data Access', 'Visualize', and 'Download dataset as ASCII'.

Data management plans

A data management plan (DMP) is a written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

Data management plans

- Are increasingly being required by funding agencies
 - e.g. RCN, EU
- And by e-infrastructure providers
 - e.g. Sigma2



File Edit View History Bookmarks Tools Help

NILU skylagring x 20200512_NorDataNet_C... x FIRST NAME's Plan x +

https://dmponline.dcc.ac.uk/plans/62809

Most Visited Getting Started NorDataNet | Arctic... XPath and XSLT with L... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva...

DMPONLINE My Dashboard Create plans Reference Help Language FIRST NAME LAST NAME

Notice: Successfully created the plan.
This plan is based on the default template.

FIRST NAME's Plan

Project Details Plan overview Write Plan Share Download

* Project title

FIRST NAME's Plan

mock project for testing, practice, or educational purposes

Funder

Grant number

e.g. 123456

Project abstract

ID

62809

Principal Investigator

Name

Select Guidance

To help you write your plan, DMPonline provides you guidance from a variety of organisations.

Select up to 6 organisations to see their guidance.

- Digital Curation Centre
- FAIRsFAIR - Fostering Fair Data Practices in Europe

Find guidance from additional organisations below

[See the full list](#)

Save

File Edit View History Bookmarks Tools Help

NILU skylagring x 20200512_NorDataNet_C... x FIRST NAME's Plan x +

https://dmponline.dcc.ac.uk/plans/62809/overview

Most Visited Getting Started NorDataNet | Arctic... XPath and XSLT with L... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva...

DMPONLINE My Dashboard Create plans Reference Help Language FIRST NAME LAST NAME

FIRST NAME's Plan

Project Details Plan overview Write Plan Share Download

DCC Template

This plan is based on the "DCC Template" template provided by Digital Curation Centre.

The default DCC template

Template version 0, published on 15 June 2020

Instructions

Write plan

The DCC default template

Data Collection

- What data will you collect or create?
- How will the data be collected or created?

Documentation and Metadata

- What documentation and metadata will accompany the data?

Ethics and Legal Compliance

- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights (IPR) issues?

Storage and Backup

- How will the data be stored and backed up during the research?
- How will you manage access and security?

Selection and Preservation

- Which data are of long-term value and should be retained, shared, and/or preserved?
- What is the long-term preservation plan for the dataset?

Data Sharing

- How will you share the data?

File Edit View History Bookmarks Tools Help

NILU skylagring 20200512_NorDataNet_Ch FIRST NAME's Plan - Write

https://dmponline.dcc.ac.uk/plans/62809/edit?phase_id=5491

Most Visited Getting Started NorDataNet | Arctic ... XPath and XSLT with I... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva...

DMP ONLINE My Dashboard Create plans Reference Help Language FIRST NAME LAST NAME

FIRST NAME's Plan

Project Details Plan overview **Write Plan** Share Download

expand all | collapse all 0/13

Data Collection (0 / 2)

What data will you collect or create?

B I [List] [Table] [Link] [Grid]

Save

Guidance **Comments**

DCC

Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

expand all | collapse all

File Edit View History Bookmarks Tools Help

NILU skylagring 20200512_NorDataNet_Ch EasyDMP

https://easydmp.sigma2.no/plan/start/

Most Visited Getting Started NorDataNet | Arctic ... XPath and XSLT with I... Parsing XML and HT... xslt.md.html DuckDuckGo — Priva...

easy.DMP Your plans Help o.godoy@met.no Log out





Create data management plans

Choose a template

Show 10 entries Search:

Template	Version	Description	
Horizon 2020	1	Simplified template based on Horizon 2020 guidelines.	Use
Horizon 2020 Expert	1	A shorter template based on Horizon 2020 that assumes knowledge of data management.	Use
Science Europe	2	Template for data management plans based on the Science Europe guidelines.	Use

Showing 1 to 3 of 3 entries Previous 1 Next

    [User Guide](#) [About](#) [Support](#) [Terms of use](#) [Privacy policy](#)