



DocEnhance

Workshop: Data Stewardship

Hanne Risan Johnsen & Helene N. Andreassen

UiT The Arctic University of Norway

KRECon, National Library of Technology, Prague, 10-11 November 2022



DocEnhance

Goal

- Inform about the DocEnhance Data Stewardship course.
- Invite you to consider how content of this course can be integrated into the PhD programme(s) in your institution.





Plan

- Presentation of the DocEnhance project and the Data Stewardship course
- Group discussion
- Wrap-up in plenum





DocEnhance

DocEnhance: Responding to the need for PhD training in transferable skills

Transferable
skills



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872483



DocEnhance

A Consortium of 19 partners



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872483



DocEnhance

The DocEnhance Platform

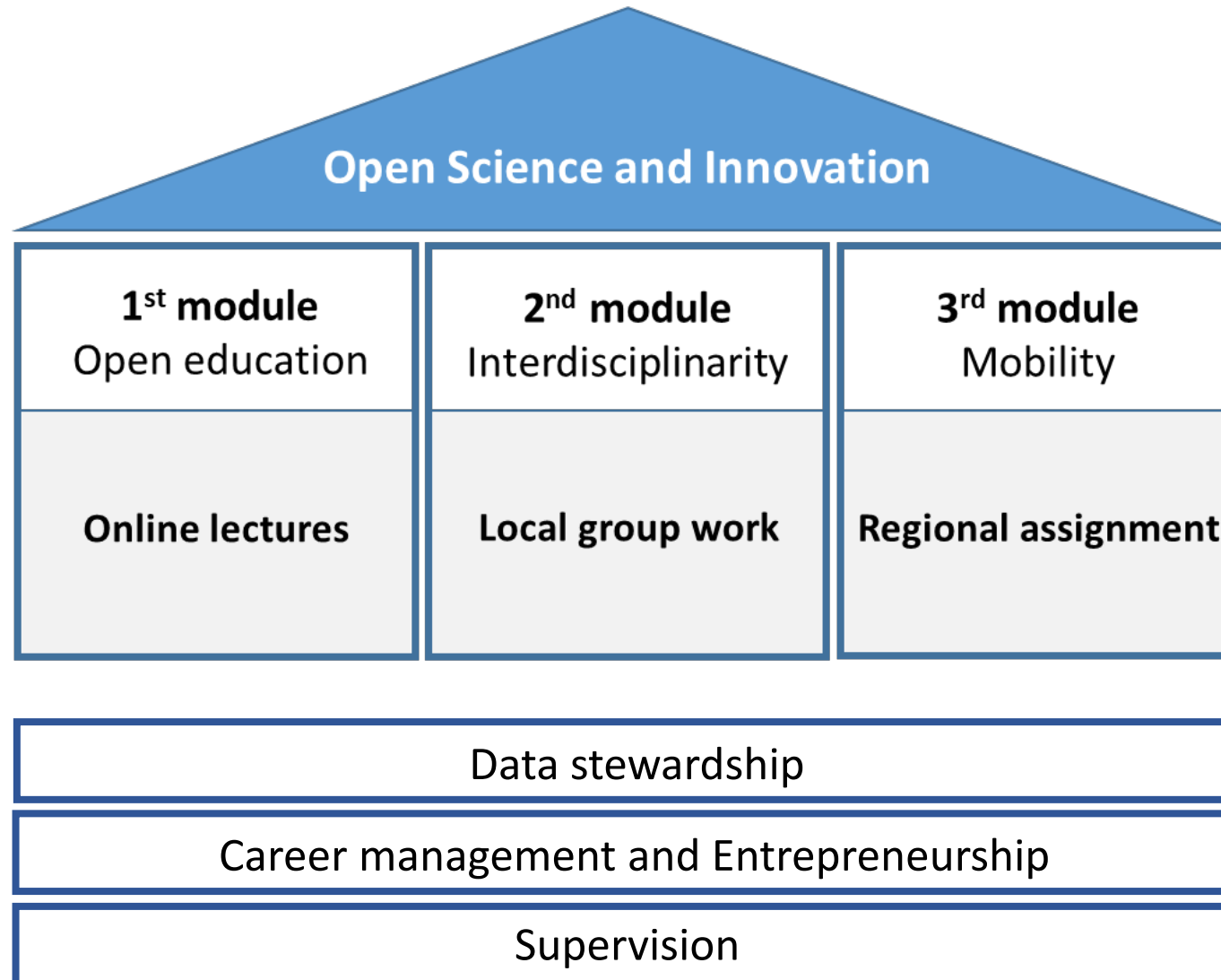


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872483



DocEnhance

Course concept





DocEnhance

The DocEnhance Platform





DocEnhance

The Data Stewardship course

Development team

Research librarians / RDM specialists /
teachers PhD education / developers OER

- Helene N. Andreassen
- Noortje Dijkstra
- Henning Hansen
- Leif Longva



Photo: Linnea Nordström



This project has received funding from the European Union's Horizon 2020 Science with and for Society programme under grant agreement no. 872483



DocEnhance

Research data management @ the UiT Library

2010- Focus on open science

2014- Focus on flexible teaching / learning

2017- Teaching & guidance on best practices
within RDM

2018- Bi-annual & multidisciplinary doctoral
course (ECTS credits) with focus on research
integrity and open science



Photo: Rune Ytreberg



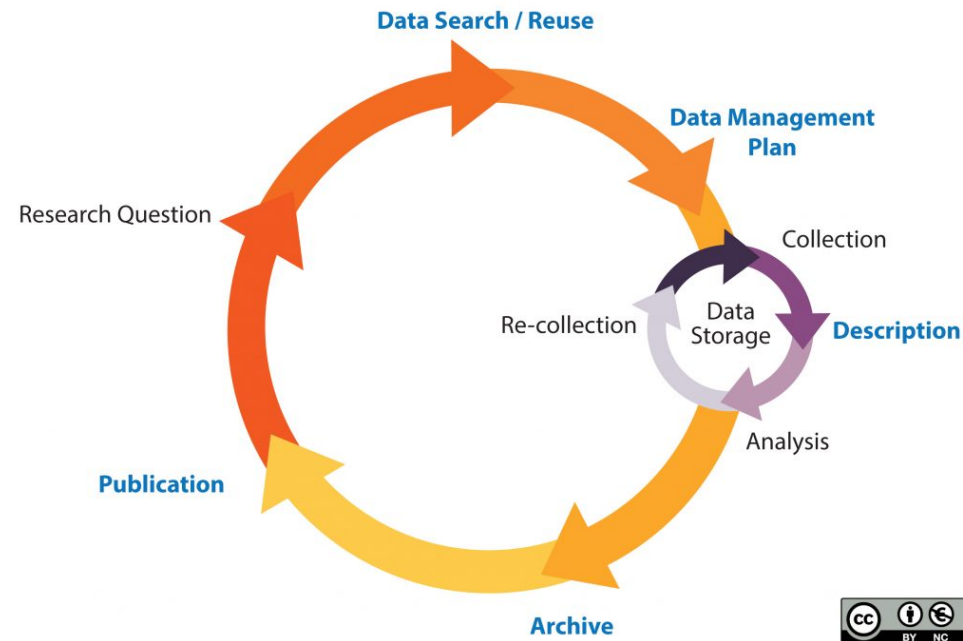
This project has received funding from the European Union's Horizon 2020 Science with and for Society programme under grant agreement no. 872483



DocEnhance

Mapping the DocEnhance course concept onto data stewardship

The Research Data Management Lifecycle



The University of California, Santa Cruz, Data Management LibGuide, Research Data Management Lifecycle. <http://guides.library.ucsc.edu/datamanagement>



This project has received funding from the European Union's Horizon 2020 Science with and for Society programme under grant agreement no. 872483



DocEnhance

Module 1

9 thematic parts

- Videos
- Food for thought
- Quizzes
- Readings

1.4 The FAIR data principles

In the [interview with Kenneth Ruud](#), there was mention of the term FAIR principles. These principles were drafted in 2014 during an international meeting and published in 2016 by Wilkinson et al.. The FAIR principles are a set of guidelines developed to optimise the reusability of research data, and they form the basis for the good research data management practices introduced in this online course. Learn more about what is behind the acronym FAIR in the following video:

The video player shows a woman with glasses and a black patterned top. To her left are two icons: one of a document with a pencil and checkmarks, and another of two people with a crossed-out circle. Below these icons are the labels 'Well documented data' and 'Clear user license'. A 'MORE VIDEOS' button is visible below the first icon. The video player interface includes a play button, volume icon, progress bar (6:22 / 7:24), and YouTube logo.

5.1 Organizing a data set

Why does it matter what names you give your data files and the folders? Well in fact, this may be crucial, in order to avoid mixing up your data and ruining your analysis. You do not want to risk that! In this video, you will learn the basics of how to organize your data in files and folders.

The video player shows a man with glasses and a dark sweater. To his left is a diagram of a folder structure with yellow and blue folders. Below the diagram is the text: '• Folders ... are helping us keep our datasets well organized and easy to browse'. A 'Watch on YouTube' button is at the bottom left. The video player interface includes a play button, volume icon, progress bar, and YouTube logo.





DocEnhance

Module 1

Multiple choice exam

- Certificate generated automatically upon successful completion
- Must be passed to be eligible for attendance in Modules 2 & 3



This project has received funding from the European Union's Horizon 2020 Science with and for Society programme under grant agreement no. 872483



Module 2

- Six three-hour long topical sessions
- Suggested exam assignment
- Participant activity and interaction in focus
- Teacher role: Selection of content, organisation and support in group/plenary discussions
- Flexibility in selection of sessions, activities, examples
 - Time available
 - Participant group (discipline, experience)
 - Preferences

For the teacher

Content

This module contains teaching material for six three-hour activity:

1. The importance of RDM and open archiving
2. The data management plan
3. Structuring and documenting research data
4. Data storage, file formats, and the FAIR principles
5. Archiving research data
6. Search, cite, and license research data

Material

Each session has the following composition:

- An overview of the session, with time frames
- Session teacher guide
- Assignment texts and other material

All material is downloadable or accessible via links. Module the PhD candidates are not supposed to enroll in the course relevant material with the PhD candidates/course participants.

Flexibility

The module is built up as an inspirational teacher's resource for a participant group and time constraints. You may also decide





DocEnhance

Module 2

Teaching material

Session 5: Archiving

Archiving data is an important step of the research data management to prepare a dataset and how to find a suitable repository for a given on datasets that are either created or modified for educational purposes.

- Structure of session 5: Archiving research data
- Structure of Session 5 (printable version)
- Teacher guide to Session 5
- 5.1 Pre-assignment text - Getting ready to work on research data archiving
- 5.2 Assignment text - Preparing and depositing a dataset
- 5.2 Dummy dataset
- 5.2 DocEnhance Educational Data Archive
- 5.2 Discussion guide for the teacher - Preparing and depositing a dataset
- 5.3 Assignment text - Selecting a research data repository

Structure of Session 5: Archiving research data

00:00 – 00:15 Plenum: Introduction to session 5

- Plan for the session.
- Recap in plenum & terminology demystifier: What do you need to think about and/or do when you want to archive research data? Teacher or someone else takes on secretary function and writes down keywords (virtual or physical whiteboard).

00:15 – 01:45 Buzz groups & plenum: Preparing and depositing a dataset

- In buzz groups, test archiving research data by entering metadata and converting/uploading files for a given dataset into the DocEnhance Educational Data Archive (for login and repository guidelines, see pre-assignment text 5.1: *Getting ready to work on research data archiving*). Write down comments, questions, or experienced challenges in a shared Padlet. Wrap-up in plenum.
- If needed, the individual buzz groups organise a 5-minute break during group work.
- Material:
 - 5.1 Pre-assignment text: *Getting ready to work on research data archiving*.
 - 5.2 Assignment text: *Preparing and depositing a dataset*.
 - 5.2 Dummy dataset.
 - 5.2 Discussion guide (for the teacher).
 - [DocEnhance Educational Data Archive](#).
 - Shared Padlet for entire group, one column per group.

01:45 – 02:00 Break

- Material: Refreshments and fresh air.

02:00 – 02:45 Buzz groups & plenum: Selecting a research data repository

- In buzz groups, select one of the datasets in the DocEnhance Educational Data Archive and search for a suitable repository. Take notes and be ready to argue for your choice. Wrap-up in plenum.
- Material:
 - 5.3 Assignment text: *Selecting a research data repository*.
 - [DocEnhance Educational Data Archive](#).

02:45 – 03:00 Sum-up and information about the next session



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101019719.



Module 3

- Advice from the non-academic sector
- Getting started
- Assignment ideas

Relevant challenges

Challenges that actors in the non-academic sector face in the context of data management: a discussion with a handful of stakeholders from the non-academic sector.

- **Data access and ownership**, e.g. how to manage data access and ownership in large organisations, procedures and collaboration with external collaborators.
- **Data protection**, e.g. how to manage data protection in large organisations, how data management is in line with regulations, how data should be managed: How much data is needed, how much knowledge is required to handle it.
- **Standardisation**, e.g. how to manage data from different entities within the organisation, how to standardise and archive the data.
- **Combination of data**, e.g. how to manage data from different datasets and their combination, how to combine specialised competences, and how to manage data.
- **Big data analysis**, e.g. how to manage data related to big data analysis, how to manage data in a way of working that rapidly is growing.
- **Data provision**, e.g. how to provide data to actors in private or public sector, how to visualise data in an understandable way.

Assignment ideas and examples

In this Module 3, the PhD candidates will work on practical assignments and work with data from the non-academic sector, while business entities will work on practical assignments and work with data from the academic sector.



Photo: "Wrapping one's head around the data" by Unhindered by Technology

Examples

Here are some possible tasks DocEnhance participants may be asked to perform:

- Analyses based on (raw) data:
 - Give the doctoral candidate a set of (raw) data from the firm/organisation and ask the doctoral candidate to:
 - Ask the doctoral candidate to suggest whether the data may be used for a specific purpose.
 - Ask the doctoral candidate to suggest how (some of the) data may be used.
 - Ask the doctoral candidate to do some basic regressions or other statistical analyses.
 - Ask the doctoral candidate to illustrate effects that these analyses may have.
 - Ask the doctoral candidate to suggest metadata to attach to the data.





Module 3

Exam assignment

- Report based on work done in collaboration with non-academic sector.



Photo: Filip Faltejsek





Feedback participants

It is important to inform early about structure, content, and materials.

Good combination of theory and application.

Great to meet and work with other PhD candidates.

“It was an awesome experience.”

“The best course I’ve done during my PhD so far.”

The course should be incorporated into the PhD study.





Feedback local teachers

Time is an issue. Enough time for each module, and enough time between modules to prepare well.

Teacher guides are essential.

Challenging the participants, getting them out of the comfort zone, is beneficial both for learning and for dynamics within the group.

The course should be given on a regular basis, and/or incorporated into the institutional PhD programme.





DocEnhance

Target groups

Individuals who want to enhance competencies and skills on RDM by self-studying.

Teachers who want elements to include in relevant teaching.

Institutions who want to include a formal course on RDM in their PhD programmes.





Implementation at the institution

- Organisation
- Teachers
- Non-academic partners

- Definition of work load
- Selection / adjustment of material
- Portefolio assessment & ECTS credits (or other)

UiT is author of the course and may be contacted for questions about content.





Group discussion

Assignment

- How can the content of Data Stewardship be integrated into PhD education and supervision at your own institution? Challenges? Possibilities?
- How can the content be made relevant and immediately useful for the different research communities? Experiences? Thoughts?
- What should be covered by such a course, and what should be covered in other contexts? Who should do what?

Keywords

- Practical issues
- Teachers
- Thematic relevance
- Rooted in the institution
- Cooperation
- Academic diversity
- PhD training programme
- PhD supervision
- Roles & division of labour





DocEnhance

Wrap-up

Share your thoughts!

Bring relevant reflections, questions and comments with you back home, and continue the discussion with your colleagues!



This project has received funding from the European Union's Horizon 2020 Science with and for Society programme under grant agreement no. 872483



DocEnhance

Course on Data Stewardship

Available by December 2022

<http://docenhance.eu>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872483