Data Stewards and Data Management Principles

(LTAT.02.014)

Priit Adler
Research Fellow, University of Tartu
Elixir EE, BIIT

2021.02.09
BIIT

Bioinformatics, Algorithmics and Data Mining Group

https://biit.cs.ut.ee/

https://elixir.ut.ee/
RDM toolkit
https://rdmkit.elixir-europe.org/
Data steward research

Keywords
Research, operational, supporting. Interacts primarily with researchers.

Description
As a research data steward, I support and work in close collaboration with the main data producers and users in academia: the researchers, ranging from undergraduate students to full professors. I advise researchers, make sure data is handled in a manner compliant with the institute's policy and may also perform hands-on work in a project.

My work focuses on implementing the institute's data guidelines and translating them into domain and project specific procedures, for example by managing a database or reviewing data management plans. My responsibilities and tasks focus on translating the researcher needs on data into infrastructural and service requirements.
Course practicalities

Lectures:
Tuesday 10.15 - 12.00, online via Zoom

Practicals / seminars:
Monday 10.15 - 12.00, online via Zoom
Course practicalities

Resources (lecture materials / homeworks):
https://courses.cs.ut.ee/2021/dsdmp/spring/

Online discussions:
https://ut-ee.zoom.us/j/91793774913
https://discord.gg/nfMEUsY2
Zoom practicalities

Due to COVID-19 all lectures and practical / seminar sessions will be held online at least until the end of March.

Be patient and take one lecture at a time.
Zoom practicalities

- Keep mic off, when not in use
- It is polite to use camera when speaking
- Use chat for comments
- Use reactions to get attention
Zoom practicalities

• Keep mic off, when not in use
• It is polite to use camera when speaking
• Use chat for comments
• Use reactions to get attention
Learning objectives

• Students can convey the importance and principles of data management to their peers
• Students know how to compile data management plan for a project
  • knowledge of the components of a data management plan
  • FAIR (Findable, Accessible, Interoperable, Reusable) principle
• Student is aware of the basic principles of data mining
• Student knows the tasks and activities of Data Steward on a project and organisation level
Grading

- Homework sums up to 50 points
- Exam is another 50 points

To pass course you need 51 points
with proviso that you need to get at least 25 points in the exam.
Desired outcome

• Data management plan *skill*
  • You are able to compile exemplary data management plan (medium bloom)
  • You are able to analyse an existing data management plan
  • You are able to assess and give feedback to an existing data management plan (high bloom)
Desired outcome

• Data steward *competence*
  • You know common tasks and activities of data steward.
  • You are able to define your role within a project or company
Training the trainers

The aim of this course is to train the trainers. i.e. you - as future data stewards

• For practical experience I plan to use reverse classroom exercises where as homework you will prepare a short topic and present it to the rest of the class

• For additional credit (substitution to exam) or a thesis topic it is possible to compile training materials in Estonian and English for data management courses.
Gentle introduction to data management buzzwords
Wiki: A **data steward** is an oversight or data governance role within an organization, and is responsible for ensuring the quality and fitness for purpose of the organization's data assets, including the metadata for those data assets.
Open Data

All EU institutions are invited to make their data publicly available whenever possible. This means data can be reused free of charge and without any copyright restrictions.

FAIR principle

Findability, Accessibility, Interoperability, and Reuse of digital assets

https://www.go-fair.org/fair-principles/
Research data lifecycle

RDM – research data management

https://rdmkit.elixir-europe.org/
7 Components of a Data Management Framework

- Data Governance: Data Governance provides the overarching support to data management through stewardship, policies, processes, standards, and adherence to leading practices.
- Data Architecture: Data Architecture provides the infrastructure for the storage, integration, and use of data throughout the organization.
- Metadata: Metadata allows you to use data more efficiently by providing critical information about data attributes.
- Data Quality: Data Quality provides the structure necessary to have data that fulfills the needs of the business.
- Data Lifecycle: The Data Lifecycle follows the data throughout the company, providing integrity from the initial introduction into the company through the final deletion from the company.
- Analytics: Analytics applies statistical and visualization techniques that lead to valuable insights that can help the company make better business decisions.
- Data Privacy: Data Privacy supports the needs of the business to share data internally and externally.

Background based biases of how we think about managing our data.

https://favogram.com/men-vs-women/
1st Homework

Self-assessment task:

Given empty DMP template, assess yourself in 5 point scale on every topic in the template:

• Do you understand what is expected of you, if ‘yes’ supplement with an example
• Are you able to help you peer to fill in given topic
• Would you be able to assess completed DMP and criticize give feedback to its owner

Submit by 15.02 23:59:59