

# **194.044 Data Stewardship (VO 2,0) 2019S** and **194.045 Data Stewardship (UE 2,0) 2019S**



Andreas Rauber, Tomasz Miksa  
TU Wien



- Management of data to ensure its
  - Availability
  - Quality
  - Interpretability
  - Reproducibility
  - ...
- Must be addressed at different levels
  - Technical
  - Organisational
  - Legal
- Data Stewardship is interdisciplinary

# Overview of Lecture Topics

- Data repositories
  - Software architecture, interfaces, interoperability, data models, standards, ontologies, etc.
  - Certification, processes, management, etc.
  - Access modes, security, privacy, etc.
- Data Management
  - Plans, policies, support services, stewards, etc.
  - Legal frameworks, roles and responsibilities, etc.
- Reproducibility
  - Software execution tracing, dependencies, containers, virtualisation, service monitoring, etc.
  - Data identification and citation, provenance, etc.
- Digital Preservation
  - migration, emulation, planning, risk management, etc.

# Lecture Goals

- As a student/researcher/employee
  - Better design and organise your work with data (and code)
    - Select proper tools and software
    - Better design data transformation processes
    - Make aware decisions on experiment/tool/process design in view of data
- As a data steward/consultant/architect
  - Have a good understanding of all components needed to establish systems/workflows/practices at an organisation

# Lecture Dates

- 11.3: Intro + Data Management Plans
- 18.3: Digital Preservation
- 25.3: OAIS Standard, Preservation Planning
- 1.4: (no lecture)
- 8.4: Data Citation
- 15.4: (no lectures: Easter break)
- 22.4: (no lectures: Easter break)
- 29.4: Machine-actionable DMPs
- 6.5: Repositories and Certification
- 20.5: Institution-wide Data Management
- 27.5: Reproducibility
- 3.6: (invited lecture - TBA)

- First exam
  - **19.06.2019, 13:00-14:00, EI9**
- Further exams
  - October 2019
  - December 2019
  - January 2020
  - March 2020
  - May 2020

- There is no single book to follow for this lecture
- Slides
  - Meant for presenting!
  - Contain links to papers/reports/webinars etc.
    - Check them!
  - Take notes
- Literature (very basic)
  - Barend Mons, Data Stewardship for Open Science: Implementing FAIR Principles, 10.1201/9781315380711  
(available online through TU Bibliothek)
  - Managing and sharing data by UK Data Archive:  
<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

# Overview of Exercises

- 194.045 Data Stewardship (UE 2,0) 2019S
  - Accompanies the lecture
  - Practical exercises to get hands-on experience
  - Programming skills essential
- Exercise 1 – FAIR Data Science (33%)
  - 11.3 – 22.4
  - Individually
- Exercise 2 – Machine-actionability and Repositories (67%)
  - 6.5 – 1.7
  - Groups of 2
- Submissions via TUWEL
- To pass
  - Min 35% of points on each exercise
  - Min 50% of point overall to be positive
- Late submission by 1 week: 10% points deduction



# Data Stewardship Activities and Collaborations

- RDA: Research Data Alliance: Data Citation, DMP Common Standards
- e-Infrastructures Austria
- European Open Science Cloud (EOSC) Secretariat (EU H2020)
- ROMOR: Research Output Management
- Timbus: Timeless Business Services (EU FP7 IP, via SBA)
- 4C: Collaboration to Clarify the Costs of Curation (FP7 CA, via SBA)
- APARSEN: (EU FP7 NoE, via SBA)
- Scape: Scalable Preservation Systems (EU FP7 IP)
- PLANETS (EU FP6 IP)
- DPE: Digital Preservation Europe (EU FP6 CA)
- ...

- Innovationslehrgang  
Data Science und  
Deep Learning
- Data Stewardship is a module

<http://www.ifs.tuwien.ac.at/idsdl/>

