RDA-P9 Joint Session of IG Data Fabric, IG Education and Training
ENVRI-RDA Summer School

Ulrich Schwardmann

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Am Fassberg, 37077 Göttingen
ulrich.schwardmann [at] gwdg.de

06 April 2017, Barcelona
Professional Data Science requires a deep understanding in techniques for efficient Data Management.

Scientific labs and also industry are desperately looking for well-trained

- Data Scientists and
- Data Managers.
goal

at the end of the course

- the participants will be able to successfully deal with a wide range of Data Science related activities.
- environmental data is used during the course,
- however, the applied methods are widely independent of the nature of the data.
will not only be of interest for
- young environmental master students,
- PhD students,
- post docs or
- experts from start-ups.

Due to the hands-on part some knowledge about programming will be necessary,
- since we will use scripting languages such as Python or PHP.
Theoretical Lectures Topics (2h)

- Mo. Data Issues in ENVRI
- Mo. Trustworthy Repositories, FAIR principles, Data Organisation, PID, MD, Certification
- Mo. D-Space Architecture, making it DFT compliant
- Tu. ENVRI metadata, existing schemas
- Tu. Metadata in D-SPACE
- Tu. PID info types, DTR intro

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Theoretical Lectures Topics (2h)

- Mo. Data Issues in ENVRI
- Mo. Trustworthy Repositories, FAIR principles, Data Organisation, PID, MD, Certification
- Mo. D-Space Architecture, making it DFT compliant
- Tu. ENVRI metadata, existing schemas
- Tu. Metadata in D-SPACE
- Tu. PID info types, DTR intro

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Practical Sessions Topics (5h)

- Mo. building a first simple repository,
- Mo. create MD + extract metadata from headers and databases,
- Mo. store data, request PIDs + create PID records, add PID to MD,
- Mo. make data accessible
- Tu. get exp data, calculate checksum,
- Tu. store exp data, create metadata, register PIDs,
- Tu. add all to repository, check accessibility, enter PID types in DTR

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Practical Sessions Topics (5h)

- Mo. building a first simple repository,
- Mo. create MD + extract metadata from headers and databases,
- Mo. store data, request PIDs + create PID records, add PID to MD,
- Mo. make data accessible
- Tu. get exp data, calculate checksum,
- Tu. store exp data, create metadata, register PIDs,
- Tu. add all to repository, check accessibility, enter PID types in DTR

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Theoretical Lectures Topics (2h)

- We. Collections, Collection Building, Citing Data
- We. Harvesting Protocols - OAI-PMH (ResourceSync)
- We. building indexes for searching, search portals
- Th. Relevance of Data Typing, defining scientific types
- Th. Designing Scientific data type entries for conversion, visualisation etc.
- Fr. Scientific Workflows
- Fr. Technical Workflows / Introducing Beaker
- Fr. R statistics in Beaker

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Theoretical Lectures Topics (2h)

- **We.** Collections, Collection Building, Citing Data
- **We.** Harvesting Protocols - OAI-PMH (ResourceSync)
- **We.** building indexes for searching, search portals
- **Th.** Relevance of Data Typing, defining scientific types
- **Th.** Designing Scientific data type entries for conversion, visualisation etc.
- **Fr.** Scientific Workflows
- **Fr.** Technical Workflows / Introducing Beaker
- **Fr.** R statistics in Beaker

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Practical Sessions Topics (5h)

- We. expose metadata to harvester B2FIND, build collections and register
- We. collections
- Th. create scientific DTR entries,
- Th. carry out visualisations, conversion
- Th. etc. incl. registration of new data
- Fr. learn Beaker, select data, write Beaker scripts to extract metadata and data,
- Fr. carry out analysis with R, create prov metadata, register new data, etc.

kind of input
- blue: user community input,
- red: data expert input,
- magenta: RDA data expert input,
Practical Sessions Topics (5h)

- We. expose metadata to harvester B2FIND, build collections and register
- We. collections
- Th. create scientific DTR entries,
- Th. carry out visualisations, conversion
- Th. etc. incl. registration of new data
- Fr. learn Beaker, select data, write Beaker scripts to extract metadata and data,
- Fr. carry out analysis with R, create prov metadata, register new data, etc.

kind of input
  - blue: user community input,
  - red: data expert input,
  - magenta: RDA data expert input,
Other tutorials

- 2015-10-07, Göttingen: Outcome of RDA P6 at PID side meeting of EUDAT2020 Kickoff
- 2015-11-25, Potsdam: PID-Tutorial at RDA-D
- 2017-03-20, Potsdam: Tutorial at RDA-D: Persistent Identifiers
- 2017-03-20, Potsdam: Tutorial at RDA-D: PID InfoTypes, DTR und Collections