Feedback from the RDA Europe - ENVRI Summer School

Marco Molinaro [INAF]

Ubi, Quando

Where, When

RDA EU-ENVRI Summerschool

on Data Management and Data Science

12-16 June 2017 Helsinki/Espoo, Finland

Learn how to do FAIR compliant professional Data Management. Experience and practice with the latest software to manage and analyse data.

CSC - IT CENTER FOR SCIENCE

RESEARCH DATA ALLIANCE

Quis

Who

- Tutors
  - Environment data managers and software specialists
  - CLARIN, ePIC experts
- Participants
  - PhD/PostDoc students (mostly)
  - Young Researchers
  - (myself)

Quid

What

- Proper Repository: how to organise and assess a proper repository with the help of a professional repository application ([CLARIN-]DSpace)
- Registering Environmental Data: upload environmental data into the repository, create metadata, assign augmented Persistent Identifiers (handle.net PIDs)
- Collection Building and Using: create collections as subject of analysis, expose metadata, cite collections, etc.
- Data Typing: use data typing as an essential element to carry out transformation, visualisation and analysis via a Data Type Registry (RDA DTR)
- Analysing Data: addressing the stored data and metadata via PIDs for analysis using BEAKER notebooks (Python and R)

Quomodo

How

- “classic” format
  - Morning presentations on topic and tools
  - Afternoon hands-on with prearranged exercises
    - Step by step guides made available
    - Mixup of
      - Leave the students alone with the exercise
      - Perform each step and wait for everybody to catch up
  - Limited to 22 participants

Quibus auxiliis

With what tools

- Starting from ready-to-use desktop machines and a set of test services
  - CLARIN-DSPACE
    - Collections repository
  - ePIC PID repository
  - Beaker server
- Python & R scripting on Beaker notebooks

Cur

Why

- School reasons
  - Data Science is considered a new thing
    - Europe is strongly pushing this role (EOSC)
  - In Europe at least it's true in the sense that there's hardly an educational path for it
    - We need to train data scientists
- My reasons for participating
  - IVOA (ASTERICS) is domain relevant
  - RDA, EOSC, other … we need to think also cross-domain
  - (TBC) there'll be a Data Science class in a Master degree course at Trieste University

Quantum

How much

(personal view on the school's impact)

- Broad overview of the interoperability scenario
- Focused on librarian and RDA cross-domain solutions
  - [IVOA] idea of service/protocol missing
  - Fields metadata annotations fine grained in repositories
- Peculiar view of data types as objects resolvable from repositories
- Hardly an idea of common data exchange formats

Conclusions

- Cross-domain Open Science and FAIR scenario is getting momentum, in Europe at least
  - EOSC projects are lacking (have poor) astronomical view (side note)
- VO is already a VRE, but we have to take care advertising this
  - RDA level is covered
  - Other projects maybe are not so aware
- Cross-domain metadata are not sufficient to cover single domain use cases
  - But then, why using repositories of data types?

Conclusions

Thank you!

- Cross-domain Open Science and FAIR scenario is getting momentum, in Europe at least
  - EOSC projects are lacking (have poor) astronomical view (side note)
- VO is already a VRE, but we have to take care advertising this
  - RDA level is covered
  - Other projects maybe are not so aware
- Cross-domain metadata are not sufficient to cover single domain use cases
  - But then, why using repositories of data types?