Engaging researchers in Research Data Management @ University of Porto

João Aguiar Castro,
INESC TEC—Faculdade de Engenharia da Universidade do Porto
TAIL Team

Cristina Ribeiro
João Rocha da Silva
João Aguiar Castro
Ricardo Amorim
Yulia Karimova
Carla Lopes
Artur Rocha
João Correia Lopes
Gabriel David
Tito Vieira
Eugénia Fernandes
Ângela Lomba
João Honrado

Infolab – Information Systems Research Group - FEUP/INESC TEC
“Second EDISON Champions Conference” - Madrid, 15/03/2017
Contents

- Research Data Management: Researchers engagement
- RDM workflow: Researchers, tasks and tools
- Curator workflow
- Case study
- Final Remarks
Research Data Management

- Funding agencies are issuing data publication mandates
- Data Management Plans as part of call regulations

**Issue (1)** Traditional publication workflow

**Issue (2)** One curator for too many domains (too many requirements)

**Issue (3)** Researchers, usually, unfamiliar with data description
Researchers engagement

Deterrents:

- additional effort that divert from main activities
- complexity of RDM tools;
- lack of awareness about the dangers of neglecting data
- belief that RDM practices are enough already
Researchers engagement

Motivators:
- credit for their work
- increased citation for datasets associated with published data
- compliance with DMP´s
- reduction of duplication – “describe once, share many times”
- streamlined communication and workflows
Researchers

Collection

Processing

Description

Publication

Data Providers

Funders

Developers

LabTablet

Dendroβ

dendro.fe.up.pt
Curator workflow

Plan → Capture → Analyse → Publish → Preserve → Reuse

Infolab – Information Systems Research Group - FEUP/INESC TEC
“Second EDISON Champions Conference” - Madrid, 15/03/2017
## Curator workflow

<table>
<thead>
<tr>
<th>Domain</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fracture Mechanics</td>
<td>Experimental</td>
</tr>
<tr>
<td>Hydrogen Production</td>
<td>Experimental</td>
</tr>
<tr>
<td>Analytical Chemistry</td>
<td>Experimental, Observational</td>
</tr>
<tr>
<td>Gravimetry</td>
<td>Experimental</td>
</tr>
<tr>
<td>Biological Oceanography</td>
<td>Observational</td>
</tr>
<tr>
<td>Biodiversity</td>
<td>Observational</td>
</tr>
<tr>
<td>Social and Behavioural Sciences</td>
<td>Observational</td>
</tr>
<tr>
<td>Cutting and Packing</td>
<td>Simulational</td>
</tr>
<tr>
<td>Vehicle Simulation</td>
<td></td>
</tr>
<tr>
<td>Computational Fluid Dynamics</td>
<td></td>
</tr>
</tbody>
</table>

Infolab – Information Systems Research Group - FEUP/INESC TEC  
“How EUDAT services support FAIR data” - Edinburgh, 20/02/2017
Curator workflow – training students

- Students from the Information Science Master’s Course at FEUP
- Data description of biodiversity datasets associated to their published papers

Experimental setup
- Training session
- Three datasets
- Three instances of Dendro: (1) DC; (2) bdv; (3) DC + bdv
- Evaluate log interactions and compare descriptions on the three scenarios
- Descriptions evaluated by a domain expert

Infolab – Information Systems Research Group - FEUP/INESC TEC
“Second EDISON Champions Conference” - Madrid, 15/03/2017
Can you provide a brief description of the datasets you are currently working with?

…the GPS data is a spreadsheet, in which the results are collected each 30 seconds….

How do you organize and described your data? Do you follow any standard?

We do not describe data, as I believe that the data are objective and simple.
Case study

Content analysis of researchers’ work

From an Experimental setup section

“One of the most critical components of a vehicle dynamics is the **driving cycle** on which all the vehicle calculations are made.”
Case study

Researchers perform data description experiments

vehicleMass ✓
nominalCapacitor ✗
Final Remarks

- Understand researchers culture. Show them RDM benefits
  - publication, citation, compliance with mandates
- Provide researchers with adequate tools may motivate them
  - DMP’s guidelines, metadata standards, data repositories…
- Comprehensive multi-disciplinary knowledge
  - licenses, preservation, analysis methods…

Stakeholders engagement is a learning opportunity for both!
Visit us

Publications and updates

http://dendro.fe.up.pt/

Dendro is *open source* on GitHub!

https://github.com/feup-infolab-rdm
Support slides
Data Management Plans

Guidelines
- H2020 on Fair Data Management
- Mantra RDM Training

Tools
- DMPonline – Digital Curation Centre
- DMPTool – University of California
Curator workflow
Related publications

- TPDL 2016
  - Usage-driven Dublin Core descriptor selection—A case study using the Dendro platform for research dataset description

- UAIS, Springer 2016
  - A comparison of research data management platforms: architecture, flexible metadata and interoperability

- iPRES 2014
  - The Dendro research data management platform: applying ontologies to long-term preservation in a collaborative environment

- Digital Libraries 2014
  - Creating lightweight ontologies for dataset description: Practical applications in a cross-domain research data management workflow
File names hold temporal information
Repositories and long-term preservation

Infolab – Information Systems Research Group - FEUP/INESC TEC
“Second EDISON Champions Conference” - Madrid, 15/03/2017