A two-stage model to reveal a university’s research data landscape and faculty’s research data practices

Thomas Seyffertitz & Michael Katzmayr

INTERNATIONAL CONFERENCE ON ECONOMICS AND BUSINESS INFORMATION, BERLIN, 6-7 MAY 2019
Academic units:
11 academic DPs (60 institutes), 15 research institutes

Research staff:
about 770 FTEs

~ 600 articles in scholarly journals per year

No institutional RDM at that time BUT RDM as a new "business" within the library

Development of RDM-concept commissioned by the vice-rector for research

Vienna University of Economics and Business Characteristics & Timeline

10/2016
• Start of systematic engagement with RDM

2017
• Empirical analysis of the research output

01/2018
• Presentation to the university management

05-11/2018
• Development of a RDM-Policy for WU

12/2018
• FDM-Policy → Legal affairs office

05/2019
• FDM-Policy becomes effective!
Motivation, Objectives & Research Design

Knowledge of the RD landscape is an important prerequisite for appropriate research data services

- **Research data landscape**: „What research data do we have at WU?“
- **Research culture**: how do researchers deal with their data; experience, data trends, needs

- **Case study design** – mixed method approach
- **Stage 1 – Document analysis** of research output → journal articles (quantitative aspects)
- **Stage 2 – Semi-structured interviews** with researchers, designed along the lifecycle of research data (qualitative aspects)
Stage 1 – Document Analysis

Analysis of Journal Articles

CRIS as source of references
Getting fulltext paper
Extract information from fulltext
Data encoding
Deriving statistical information


1. We define R-Daten.
2. R-C-Finance
3. Dererializer
4. Iris
Expert interviews can be very effective

Awareness of certain topics can be increased among the interview partners

Objective: exploring expert knowledge (in terms of technical and process knowledge)

Sample building criteria
  - Experience with data driven research
  - Covering all departments
  - Junior- and senior-researchers included

Interview/Topic-guide: designed along the researcher’s day-to-day research work and lifecycle of research data

Pre-test & 25 interviews
### Results – Analysis of the Journal Articles

#### General

- Analysed 596 articles published in 2016
- ~80% without external/third-party funding
- Only 12% funding of papers without RD
- Whereas >33% of articles containing RD received funding
- 86% quantitative RD
- Almost 30% contained both quantitative and qualitative RD

<table>
<thead>
<tr>
<th>Attributes</th>
<th>WU-level</th>
<th>SoWI-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>abs.</td>
<td>rel.</td>
</tr>
<tr>
<td><strong>Research data</strong></td>
<td>250</td>
<td>41.95%</td>
</tr>
<tr>
<td><strong>No research data</strong></td>
<td>346</td>
<td>58.05%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>596</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Funder</strong></td>
<td>126</td>
<td>21.14%</td>
</tr>
<tr>
<td><strong>No funding</strong></td>
<td>470</td>
<td>78.86%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>596</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data form</th>
<th>WU-level</th>
<th>SoWi-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>abs.</td>
<td>rel.</td>
</tr>
<tr>
<td><strong>Quantitative RD only</strong></td>
<td>141</td>
<td>56.40%</td>
</tr>
<tr>
<td><strong>Qualitative RD only</strong></td>
<td>35</td>
<td>14.00%</td>
</tr>
<tr>
<td><strong>Both forms</strong></td>
<td>74</td>
<td>29.60%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>250</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Results – Analysis of the Journal Articles

What the data are about

<table>
<thead>
<tr>
<th>Types of data</th>
<th>WU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic data</td>
<td>65</td>
</tr>
<tr>
<td>Company data</td>
<td>56</td>
</tr>
<tr>
<td>Social research data</td>
<td>97</td>
</tr>
<tr>
<td>Technical systems data</td>
<td>11</td>
</tr>
<tr>
<td>Environmental and natural science data</td>
<td>15</td>
</tr>
<tr>
<td>Other</td>
<td>75</td>
</tr>
<tr>
<td>Total frequency over all articles with RD (n=250, WU)</td>
<td>319</td>
</tr>
</tbody>
</table>

Percentage of different types of data occurring in articles:

- Economic data: 20.38%
- Company data: 17.55%
- Social research data: 30.41%
- Technical systems data: 3.45%
- Environmental and natural science data: 4.70%
- Other data: 23.51%
Results - Analysis of Journal Articles

Data Format Types

<table>
<thead>
<tr>
<th>Data format types</th>
<th>WU (abs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>9</td>
</tr>
<tr>
<td>Audio</td>
<td>33</td>
</tr>
<tr>
<td>Video</td>
<td>4</td>
</tr>
<tr>
<td>Alphanumeric data</td>
<td>249</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>Total frequency over all articles with RD (n=250)</td>
<td>296</td>
</tr>
</tbody>
</table>

Percentage of data format types occurring in articles:
- Image: 3.04%
- Audio: 11.15%
- Video: 1.35%
- Other: 0.34%
- Alphanumeric data: 84.12%
Analysing the Interviews
Following Meuser and Nagel (1991, 2009)

- **Transcription**
- **Paraphrase** - sequencing of the text according to thematic units
- **Coding** – ordering paraphrased passages thematically along our topic-guide
- **Thematic comparison** – grouping comparable passages from interviews
- **Conceptualization** - generalizing restricted to the empirical data
- **Theoretical generalization**

Analysis of a single interview

Analysis across multiple interviews
**Semi-structured Interviews - Findings**

**DMP, research funders**
- Some experience existing with funder mandates
- DMPs have been used for projects funded by DFG, ERC, ESRC and Horizon 2020

**Data trends and developments**
- Quantitative research methods: Big Data
- Source: increasingly WWW, Social media
- Storage; computer performance

**Managing RD within the research process**
- Data are scattered, data management strategy within departments rather an exception
- Use of external cloud-systems

**RD in the publication process**
- Quantitative RD: Publishers’ data-policies relevant
- Qualitative RD: data usually not published; no policies (e.g. sociology)
<table>
<thead>
<tr>
<th>Section</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archiving RD</td>
<td>- In a pragmatic and case-based way</td>
</tr>
<tr>
<td></td>
<td>- effort required for the accurate description of RD</td>
</tr>
<tr>
<td>Data loss</td>
<td>- Respondents have already experienced data loss</td>
</tr>
<tr>
<td></td>
<td>- Related to processed RD</td>
</tr>
<tr>
<td>Sharing &amp; Reuse</td>
<td>- Most regard reuse as relevant (reproducibility)</td>
</tr>
<tr>
<td></td>
<td>- Sometimes data „used up“; dependent on discipline</td>
</tr>
<tr>
<td>Research data-Policy</td>
<td>- Different (discipline-specific) cultures</td>
</tr>
<tr>
<td></td>
<td>- Some fear over-regulation</td>
</tr>
<tr>
<td>RDM as a service</td>
<td>- Awareness-raising measures, one-stop shop</td>
</tr>
<tr>
<td></td>
<td>- Need for advice; information &amp; technical services</td>
</tr>
</tbody>
</table>
Recommendations Based on Our Findings

1. **Developing a RDM-Policy**: awareness, research cultures; optional implementation at department level

2. **Online material at the university’s website**: summarizing relevant information

3. **Single point of contact**: providing information on RDM-services, referring to other sources etc...
Developing a RDM-Policy for WU
Several Important Aspects

Organisational and content-related issues
- Single policy for the entire university or framework with optional customization at the academic department
- Recommending or more regulating character

Basic orientation when developing a first draft
- Based on state-of-the-art examples
- Taking into account empirical findings and the research cultures at WU
- Middle ground between directive specifications and non-binding recommendations
- The draft defines an ethical standard in dealing with research data and is therefore deliberately designed as a policy and not as guideline
Selected Literature

Methodology


Conceptualising data

Thank you for your attention! Questions?