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

Vienna University of Economics and Business

Characteristics & Timeline

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

10/2016
- Start of systematic engagement with RDM

2017
- Empirical analysis of the research output

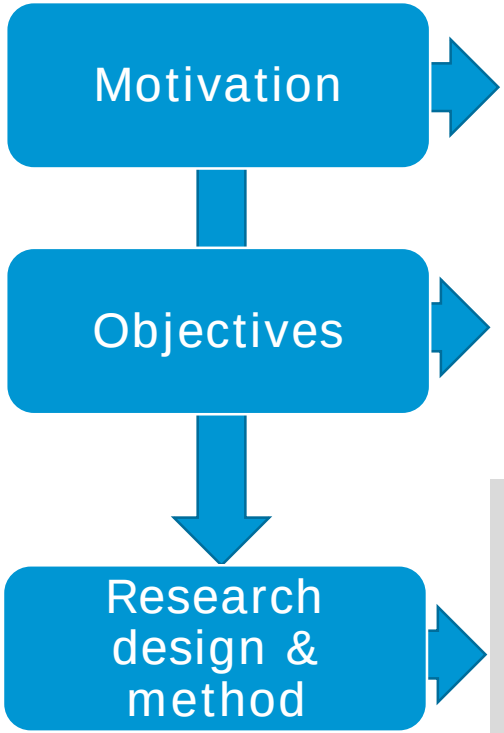
01/2018
- Presentation to the university management

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



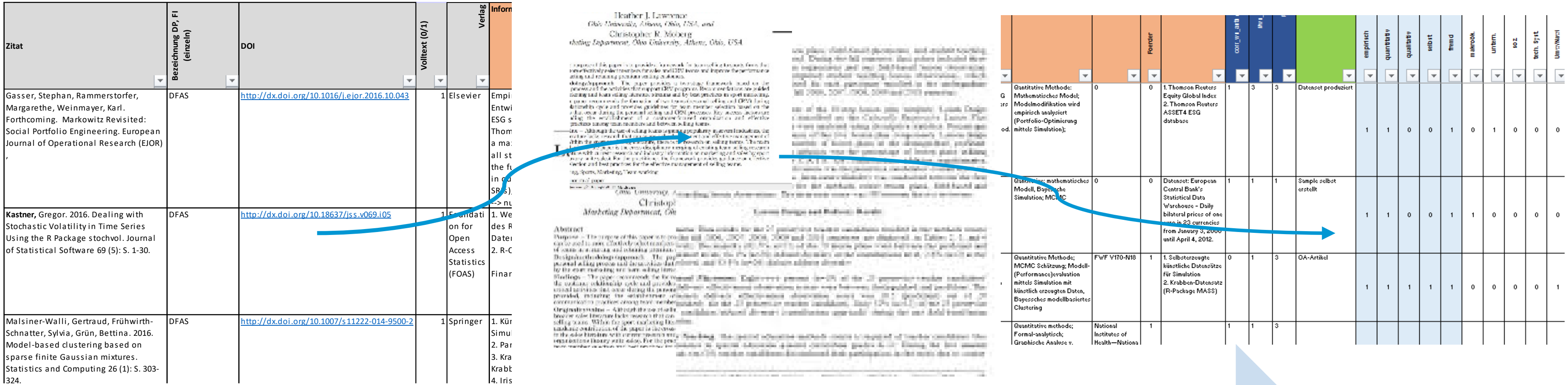
Knowledge of the **RD landscape** is an important prerequisite for appropriate research data services

- **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



1. CRIS as source of references
2. Getting fulltext paper
3. Extract information from fulltext
4. Data encoding
5. Deriving statistical information

Stage 2 – Semi-structured Interviews

Rationale & Methodological Aspects

- 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

Attributes	WU-level		SoWI-level	
	abs.	rel.	abs.	rel.
Research data	**250**	41,95%	**247**	57,18%
No research data	**346**	58,05%	**185**	42,82%
Total	**596**	100,00%	**432**	100,00%
Funder	126	21,14%	110	25,46%
No funding	470	78,86%	322	74,54%
Total	**596**	100,00%	**432**	100,00%
Data form	WU-level		SoWi-level	
	abs.	rel.	abs.	rel.
Quantitative RD only	141	56,40%	139	56,28%
Qualitative RD only	35	14,00%	34	13,77%
Both forms	74	29,60%	74	29,96%
Total	**250**	100,00%	**247**	100,00%

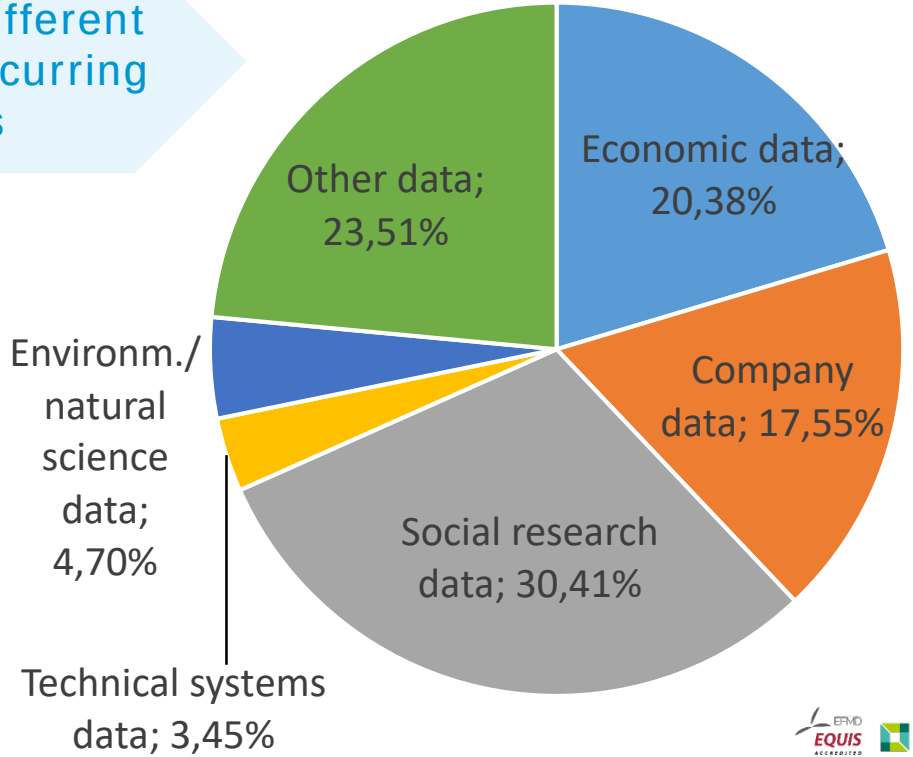
- Analysed 596 articles published in 2016
- ~80% without external/third-party funding
- Only 12% funding of papers without RD
- Whereas >33% of articles containing RD received funding
- 86% quantitative RD
- Almost 30% contained both quantitative and qualitative RD

Results – Analysis of the Journal Articles

What the data are about

Percentage of different types of data occurring in articles

Types of data	WU
Economic data	65
Company data	56
Social research data	97
Technical systems data	11
Environmental and natural science data	15
Other	75
Total frequency over all articles with RD (n=250, WU)	319

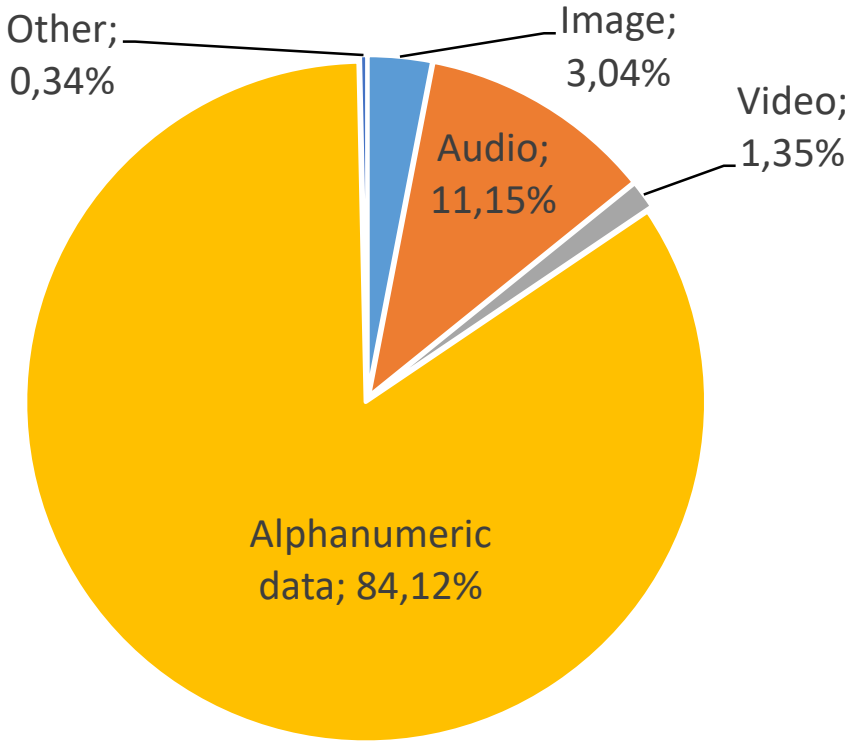


Results - Analysis of Journal Articles

Data Format Types

Percentage of data format types occurring in articles

Data format types	WU (abs.)
Image	9
Audio	33
Video	4
Alphanumeric data	249
Other	1
Total frequency over all articles with RD (n=250)	296



Analysing the Interviews

Following Meuser and Nagel (1991, 2009)

– Transcription

Condensing

– **Paraphrase** - sequencing of the text according to thematic units

– **Coding** – ordering paraphrased passages thematically along our topic-guide

Analysis of a *single* interview

– **Thematic comparison** – grouping comparable passages from interviews

– **Conceptualization** - generalizing restricted to the empirical data

Analysis across *multiple* interviews

– Theoretical generalization

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)

Semi-structured Interviews – Findings ctd.

Archiving RD	• In a pragmatic and case-based way • effort required for the accurate description of RD
Data loss	• Respondents have already experienced data loss • Related to processed RD
Sharing & Reuse	• Most regard reuse as relevant (reproducibility) • Sometimes data „used up“; dependent on discipline
Research data-Policy	• Different (discipline-specific) cultures • Some fear over-regulation
RDM as a service	• Awareness-raising measures, one-stop shop • Need for advice; information & technical services

Recommendations Based on Our Findings

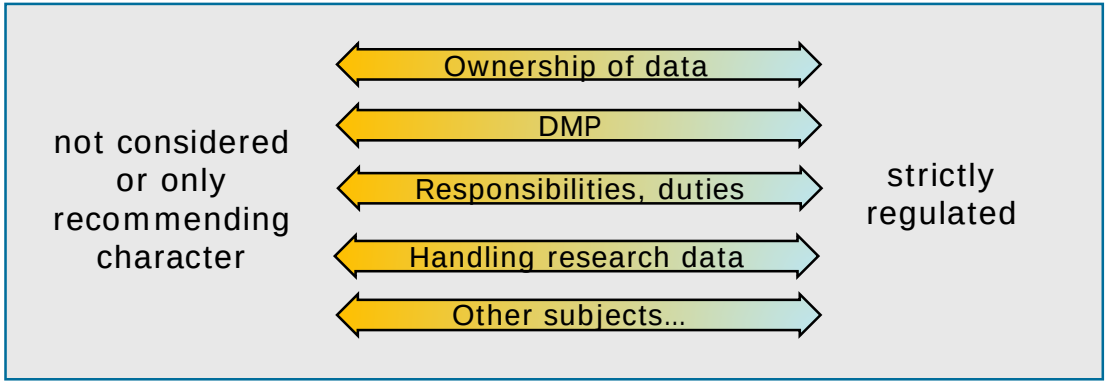
1. **Developing a RDM-Policy**: awarness, research cultures; optional implementation at department level
2. **Online material at the university‘s website** summarizing relevent information
3. **Single point of contact**: providing information on RDM-services, referring to other sources etc...

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

- Bogner A., Menz W (2009): The Theory-Generating Expert Interview. Epistemological Interest, Forms of Knowledge, Interaction. In: A. Bogner, B. Littig and W. Menz (eds.): *Interviewing Experts*. London [UK], Palgrave Macmillan, pp. 43–80.
- Meuser M., Nagel U. (2009): The Expert Interview and Changes in Knowledge Production. In: A. Bogner, B. Littig and W. Menz (eds.): *Interviewing Experts*. London [UK], Palgrave Macmillan, pp. 17–42.
- Pickard J., Childs S (2013): *Research methods in information*. London, Facet.
- Yin R. K. (2017): *Case study research and applications - design and methods*. London, SAGE.

Conceptualising data

- Kitchin, R. (2014): The data revolution: big data, open data, data infrastructures & their consequences. Thousand Oaks [CA], SAGE.

Thank you for your attention!

Questions?