Universität
Konstanz

Research Data Management
at 9 Universities in Baden-Wuerttemberg, Germany
The Results from the Final Report
of the bwFDM Communities Project

Karlheinz Pappenberger

Tromsø, 10th Munin Conference on Scholarly Publishing, 30/11/2015

The e-science initiative in Baden-Wuerttemberg

9 universities
30,000 academic staff
176,000 students

The e-science initiative in Baden-Wuerttemberg

The concept paper (published 2014)

- *E-Science – Science in a New Environment. Further development of the scientific infrastructure in Baden-Wuerttemberg.* Ministry of Science, Research and the Arts (ed.)

5 defined action areas

- Licensing electronic information media
- Digitisation
- Open access
- **Research data management**
- Virtual research environments

➡ working out action plans / initiating projects / developing products and services

bwFDM communities:
a federal research data management project

bwFDM = Baden-Wuerttemberg Forschungsdatenmanagement
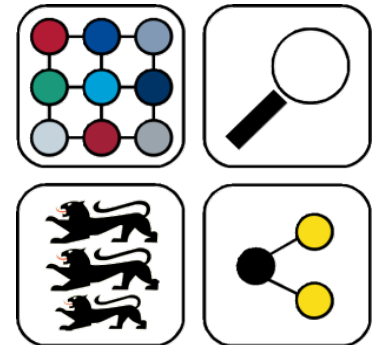**(Baden-Wuerttemberg research data management)**

Funding

- Funder: Ministry of Science, Research and the Arts
  Baden-Wuerttemberg
- Funding period: 01/2014 – 06/2015

Staff

- 9 full time key accounters (one at each university)
- 1 project coordinator

Project goal

➢ Executing a comprehensive (qualitative) survey among researchers

bwFDM communities:
milestones of the survey

List of research groups and contact persons (2/2014-4/2014; n ~ 3,000)

Guideline for interviews („how are you working?" / „what are your needs?"; ~ 40 questions)

Interviewing researchers and transforming the results into a machine readable format (2/2014-11/2014: 627 interviews with 779 researchers; in average one hour / interview)

Extracting "user stories" / statements and collecting them in a database (11/2014-2/2015: 2,554 user stories: humanities 795 / science 467 / life science 703 / engineering 589)

Compiling thematic areas and building working groups on them (2/2015-6/2015)

Grouping user stories in story maps (2/2015-6/2015; 18 topics)

Report / presentation of results and recommendations to the ministry (17th July 2015 presentation; official report probably Dec. 2015; summary in English)
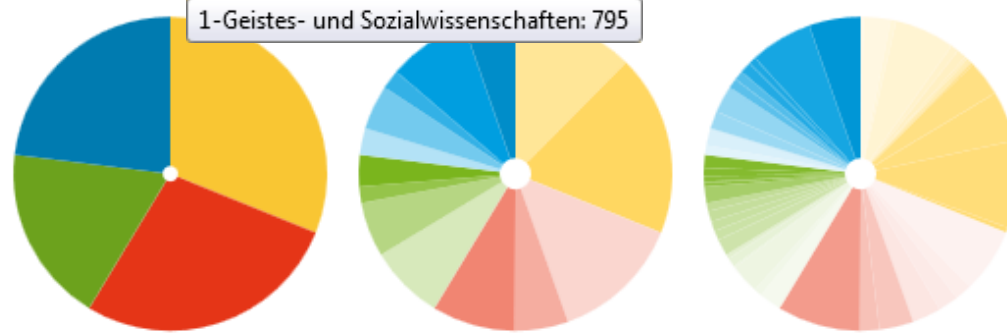
bwFDM communities:
user stories – two examples

"As researchers, who work a lot with personal data, we desire to simplify privacy policies, to outsource the dealing with data protection outside research groups and to implement a checklist on data protection in order to make the handling of personal data in our research easier and more efficient."

"As researchers in a large institute that accumulated data for over 100 years, we want FDM knowledge, in particular how to deal with legacy data (store, recycle or throw away?) and how to document and prepare data reusable for others to leave a value for future generations (even if, for example, the database software changes)."

bwFDM communities:
the 2,554 user stories (http://bwfdm.scc.kit.edu/daten/)

User Storys

Disziplinen nach DFG-Fachsystematik (interaktiv!)

1-Geistes- und Sozialwissenschaften: 795

Themenkategorien im Abschlussbericht

2-Rahmenbedingungen
3-Verarbeitung
4-Erhebung und -austausch
5-Bewahrung
6-Weitere Aspekte

0   500

Die Farben und Bezeichnungen (Mausüberfahrt) der DFG-Systematik stammen von hier.

Schlagwortsortierung der Tabelle [Try it]
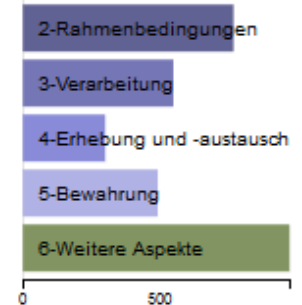
2.1-Recht,4.3-Zusammenarbeit,6.1-IT-Infrastruktur	Als Forschende, die viel mit personenbezogenen Daten arbeiten, wünschen wir uns einen zentralen Cloudservice, der es erlaubt personenbezogene Daten abzuspeichern, um keine eigene OwnCloud-Instanz zu betreiben und aufwendig administrieren zu müssen, um personenbezogene Daten abspeichern, aber auch Daten mit anderen austauschen zu können.
2.1-Recht,5.3-Archivierung	Als Forschende, die mit sensiblen Daten arbeiten sowie eine langfristige Datenspeicherung für wichtig halten, wünschen wir uns einen sicheren, datenschutzkonformen Archivierungsdienst, um sensible Forschungsdaten nach DFG-Richtlinien langfristig und sicher sowie mit einer Zugriffsmöglichkeit für uns und unsere Kooperationspartner archivieren zu können.
2.1-Recht,5.3-Archivierung	Als Forschende, die eigene Daten langfristig archivieren möchten, wünschen wir uns eine grundsätzliche einfache Archivierungslösung, mit klaren Richtlinien betreffend der Vorgehensweise (sowohl Form als auch Datenschutzkonformität), um unsere personenbezogenen Forschungsdaten aus Projekten nach DFG-Richtlinie 10 Jahre korrekt und langfristig speichern zu können.
2.1-Recht,5.3-Archivierung	Als Forschende, die empirisch forschen, wünschen wir uns eine zentrale Möglichkeit, unsere personenbezogenen Forschungsdaten in einer Größenordnung von 100 GB pro Jahr langfristig archivieren zu können, um diese bei Bedarf für die eigene Forschung nachnutzen zu können.

bwFDM communities: the results

A) General requirements and policy framework

(1) Information services

− many of the interviewed groups don´t feel well-informed about RDM

Advice and expertise in RDM
− Newsletter (e.g. university research support)
− Information platform
− Information centre / contact person
− Training / tutorials
− RDM guideline / RDM policy
− RDM teaching courses
− General information requests in RDM

bwFDM communities: the results

A) General requirements and policy framework

(2) Legal Issues: intellectual property rights, copyright and data protection

Advice and expertise in IPR and data protection / publishing data
− Information centre / contact person

Fewer legal restrictions in IPR and data protection
− When using data from others

Stronger property rights
− For own collected and processed data

IT infrastructure that fulfills data protection requirements
− When sharing data (within a project group)
− When archiving data

bwFDM communities: the results

A) General requirements and policy framework

(3) Scientific culture on data

- half are satisfied with the availability of data, half are not
- half have data that might be of interest to others but don´t share

Incentives

Status quo: Limited exchange of data in scientific communities
- Time
- Personal risk / career

- ➢ To be solved within the scientific community and the funding organizations

bwFDM communities: the results

B) Technical framework for more effective data processing

(4) Standards and formats

- Software
- Types of files, exchange formats

(5) Discipline specific and general IT-support

- Articulated often in the area of digital humanties
- Data compression
- Image processing and image recognition
- Support in statistical analysis
- Text analysis
- Simulation
- Visualization tools

bwFDM communities: the results

C) Data collection and data sharing

(6) Access to commercial and governmental data / data of NGOs

(7) Digitization

- Articulated mainly in humanities
- Using digitized material / active digitization of material

(8) Scientific cooperation

- Management and exchange of data unsatisfactory at the moment:
  (often named: email / USB-stick / dropbox / server / …)

- Virtual research environments

bwFDM communities: the results

D) IT infrastructure / IT support

(9) Storage
- Very often named / Efficient access (speed, simplicity) / Archiving: 10 years +

(10) Computing power / high performance computing needs

(11) Hardware
- Special requirements
- Easier and more flexible purchasing / Money

(12) Software / Software tools
- Access to specialist software
- (campus) software licensing

(13) IT support
- for using IT infrastructure / for data processing
- [more IT staff at the research group level]

bwFDM communities: the results

E) Preservation

(14) Documentation of projects and data

Project documentation / documentation of data
- RDM plan and guidelines
- RDM information centre / Accompanying consulting by a RDM expert
- Data curation / Support for data curation
- Research information system for documentation

Metadata
- Metadata standards
- Professional staff for data enrichment with metadata / automation

bwFDM communities: the results

E) Preservation

(15) Data repositories

- Central / structured / curated
- Both disciplinary and interdisciplinary
- Definable access rules
- Visualization of data

Finding data / Using data of others

- General search engine for data
- Access to high-quality data

(16) Archiving

bwFDM communities: the results

F) Remaining topics

17) Funding / financial issues

- More money at the research group level

18) Open Science / Open Data / Open Access

- Open source software
- Data curation
- Trust

Concluding remarks

➢ **Open project data:** huge and unique dataset probably with relevance also outside Baden-Wuerttemberg: Open data: http://bwfdm.scc.kit.edu/daten/ (with all 2,554 user stories)

➢ **Heterogeneous demands:** IT driven project, but:
many of the answers cover less technical needs than expected
and are much more heterogeneous than expected

➢ **RDM infrastructure support has to be convincing:**
Researchers also with reservations about RDM:
    Efficiency, special needs, bureaucracy, time pressure

➢ **Infrastructural competence and cooperation:**
Several players: computing centres, data centres, libraries…
    with a lot of distributed expertise → establishing appropriate services and tools
Thereby: Cooperation within an institution and also on a state/national/international level

➢ **BW strategy**: 11/2015: 8 new RDM projects:1 RDM info-platform / 7 IT; up to 3 years funding

Universität
Konstanz
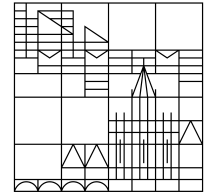
Thank you
Very much!

Karlheinz Pappenberger

Subject Librarian for economics and statistics

Specialist for research data management

University of Konstanz - KIM

karlheinz.pappenberger@uni-konstanz.de

Project: http://bwfdm.scc.kit.edu/english/

http://bwfdm.scc.kit.edu/daten/