

# bwFDM Communities

a Research Data Management

Initiative in the State of

Baden-Wuerttemberg



Karlheinz Pappenberger

Tromsø, 9th Munin Conference on Scholarly Publishing, 27/11/2014

# Overview

## 1) Federalism in Germany: Baden-Wuerttemberg and its universities

## 2) The e-science initiative in Baden-Wuerttemberg

## 3) bwFDM Communities: a research data management project

- the family of „bw“-projects
- funding
- purpose
- milestones
- results

## 4) Research data management at the University of Konstanz

- Movebank Data Repository
- developing expertise and support

## 5) Concluding remarks

# 1) Baden-Wuerttemberg and its 9 universities



## 9 universities in numbers

Universities in Baden-Wuerttemberg	Academic Staff (all: 30,300)	Students (all: 176,000)
Heidelberg*	5,500	31,500
Tübingen*	4,450	27,200
Freiburg	6,800	24,700
Stuttgart	3,350	24,600
Karlsruhe	6,000	24,500
Mannheim	1,000	12,300
Konstanz*	1,250	11,800
Hohenheim	950	9,900
Ulm	1,050	9,500

\* = member of the group of 11 German universities with nationally funded institutional strategies within the German top-level research initiative

## 2) The e-science initiative in Baden-Wuerttemberg (I)

### the concept paper

- *E-Science – Science in a New Environment. Further development of the scientific infrastructure in Baden-Württemberg.*
- published 29<sup>th</sup> July 2014, 120 pages (only available in German)
- edited by the Ministry of Science, Research and the Arts
- 3.7 million euro for working out action plans

### 5 defined action areas

- licensing electronic information media
- digitisation
- open access
- research data management
- virtual research environments

## 2) The e-science initiative in Baden-Wuerttemberg (II)

### the concept paper – chapter 4: research data management

- written by a working group of 10 members, lead by the ministry and with representatives from
  - computing centres
  - data centres
  - libraries
  - researcher communities
- recommendations
  - further development of technical „bw“-infrastructure
  - establishing research data management within teaching and curricula
  - countrywide research data repository
  - data life cycle labs
  - coordination of policies, legal issues and standardisation questions

### 3) bwFDM Communities: a federal research data management project

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

**as part of several existing and planned IT-driven „bw“-tools:**

- bwHPC: high-performance computing
  - bwSync&Share: joint sharing and synchronisation of data
  - bwFileStorage: additional countrywide storage capacity
  - bwIDM: identity management
  - bwMS: licensing of Microsoft software
  
  - bwCloud\*: federal virtualisation of server and IT-services
  - bwDataArchiv\*: long term preservation of data
  - bwDataDiss\*: research data of dissertations
  - ....
- \* = project status

# bwFDM communities: the quest for effective research support

## background: the data landscape

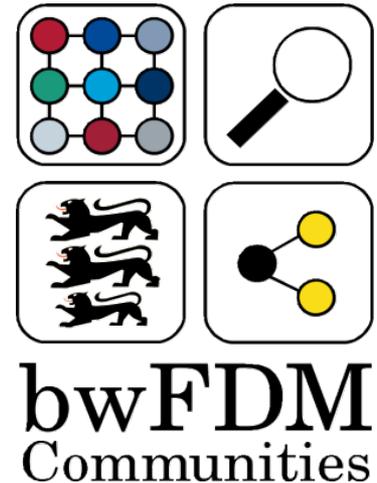
- data life cycle
- complex and large scale data
  - relevant for a growing number of science disciplines
- intelligent data management and analysis creates more high-quality research and publications



# bwFDM communities – the project

## funding

- funder: Ministry of Science, Research and the Arts  
Baden-Wuerttemberg
- funding period: 2014/01 – 2015/06
- budget: 1 million euro
- staff: 9 full time key accounters (one at each university)  
1 project coordinator
- project lead: Steinbuch Computing Centre Karlsruhe Institute of Technology
  - main project partners: computing centres at the 9 universities
  - associated: libraries



# **„What infrastructures and services are needed to make the region a globally leading area in research and development?“**

improving scientific research in Baden-Wuerttemberg in context of the e-science strategy

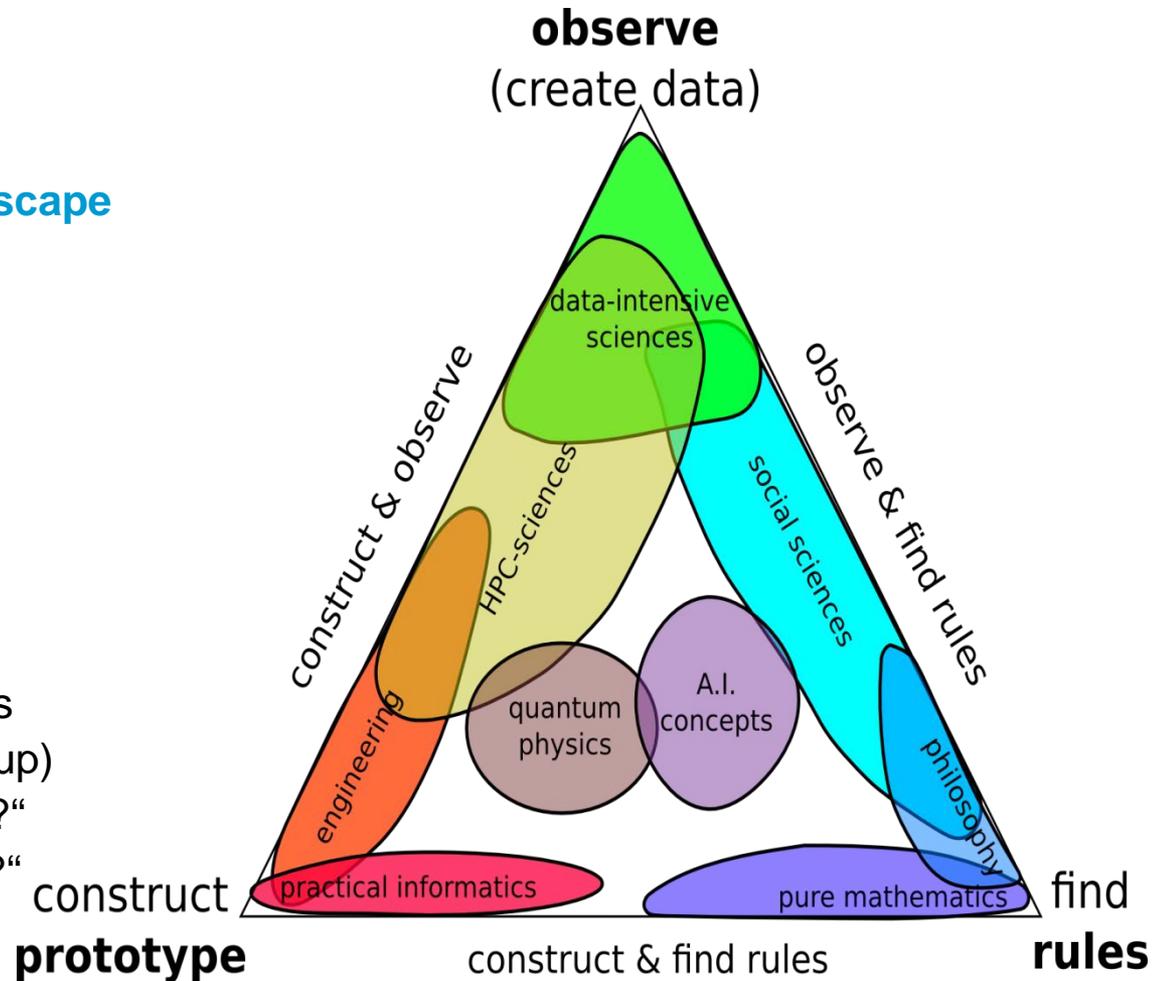
main question & project goal of bwFDM: how to do this?  
→ executing a comprehensive survey

- detailed recommendations for concrete steps
- definition of future projects and tasks

# bwFDM: the research communities and the survey

## the observed research data landscape in Baden-Wuerttemberg

- 3000 different research groups
  - applied sciences
  - life science incl. medicine
  - social sciences
  - humanities
- 700 interviews with researchers (approx. 1h each research group)
  - „how are you working?“
  - „what are your needs?“
- >1600 user stories extracted from the interviews



# bwFDM: milestones of the survey

List of research groups and **contact persons** (from Feb 2014 on)

**Guideline** for interviews

**Interviewing researchers** and transforming the results into a machine readable format (Feb 2014 to Nov 2014)

Extracting „**user stories**“ / statements and collecting them in a database

Compiling **thematic areas** and building working groups on them

Grouping user stories in **story maps**

**Report** / presentation of results and recommendations to the ministry (30th June 2015, hopefully available in English too → sharing information)

ID\*  
1114

Modifiziert\*  
2014-11-26 08:47:55

Standort  


Fachbereich  


Status

Example of an extracted userstory (1 of 1635)

Rollenfloskel  
Als

Hintergrund  
Wissenschaftler, der oft interessante Themen aus Datenschutzgründen nicht verfolgen kann,

Sharing

Processing

Erhebung

Archiv

OpenAccess

Personenservices

MetaDaten

Grundwissen

Web

Rahmen

Wunschfloskel  
wünsche ich mir

Wunsch  
eine Lockerung der Bestimmungen zum Datenschutz und weniger Einschränkungen insb. bei amtlicher Statistik

Sonstiges

Allgemein

Datenfern

Hilfe

Fertig

Zweckfloskel  
um

Zweck  
die für unsere Forschung nötige Verfügbarkeit kleinräumig zugeordneter und deshalb nicht anonymisierbarer Daten (z.B. Daten zu Wanderungsbewegungen der Bevölkerung, Wirtschaftsdaten) zu ermöglichen.

Wehrle

vanEkeris

Kratzke

Shcherbakov

Tristram

Cayoglu

Rex

Knopp

Hertzer

Svoboda

Schwabe

Bamberger

Anmerkung  
Mögliche Themen werden wegen Datenschutz oft nicht verfolgt, in der Folge unterbleibt Forschung daran (z.B. Daten zu Wanderungsbewegungen, Wirtschaftsdaten). Das Fachgebiet erfordert kleinräumige Zuordnung, dies widerspricht einer Anonymisierung.  
Welche Einschränkungen sind hier gemeint? (PB)  
Dass er gewünschte Daten nicht bekommt, weil diese nicht anonymisiert werden können. (He)

## bwFDM: most frequent user story contents

- storage
- more staff
- money for licensing commercial data, for software on analyzing data...
- money to publish open access / open data
- information about new services, tools, rules...
- data sharing
- open access data repository
- data repository with defined access rights
- assistance for legal questions / property rights on data...
- assistance for analyzing data
- assistance / standards in writing RDM plans and reports
- tools for managing one's own data
- better access to data of others / to official/governmental data
- help how to find data
- standards when working with and sharing data
- long term preservation
- digitization of data

# bwFDM: first results

## technical infrastructure

- of course: expressed demand in developing tools and services, but:
  - more heterogeneous than expected
  - fraction of answers covers less technical and infrastructural needs than expected

## beyond technical infrastructure

- many requests for a better work environment to make research easier!
  - clarification of requirements and existing know-how (RDM plans...)
  - proclamation of support, facilities, opportunities
  - standardized formats, methods, services
  - simplification of processes (fair use/dealing)
  - methodological assistance
  - long-term perspectives

→ Wide range of existing knowledge and requested needs on data

→ What to do with the results?

## 4) Research data management at the University of Konstanz - an example of a local approach

### A first project: Movebank Data Repository (topic: animal tracking)

#### funding

- German Research Foundation (DFG) 10/2010 – 9/2012
- part of a broader „Movebank Virtual Research Environment“ project
  - *Max Planck Institute for Ornithology*: Movebank Database
  - *Informatics*: visualization and data analysis
  - *Computing Centre of the University of Konstanz*: collaborative tools
  - *Library of the University of Konstanz*: data repository

#### results

- discipline specific data repository online from 2012 (based on dSpace)
- published datasets in combination with published articles
  - 24 data packages (11/2014) / 50 data files (.csv / .txt)
  - doi as persistent identifier
  - indexed by Thomson Reuters Data Citation Index



## Welcome to Movebank!

Welcome to Movebank! Movebank is a free, online database of animal tracking data hosted by the **Max Planck Institute for Ornithology**. We help animal tracking researchers to manage, share, protect, analyze, and archive their data. Movebank is an international project with over four thousand users, including people from research and conservation groups around the world.

**How does Movebank work?** The animal tracking data accessible through Movebank belongs to researchers all over the world. These researchers can choose to make part or all of their study information and animal tracks visible to other registered users, or to the public.

[Browse existing tracks on Movebank](#)

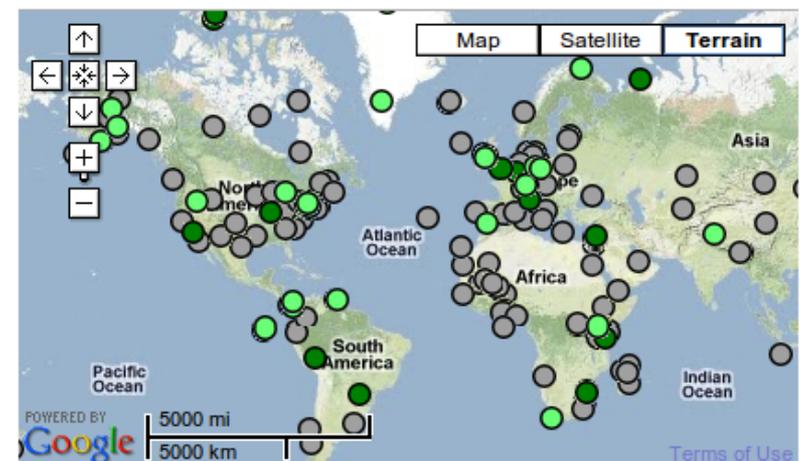
[Add data to Movebank](#)

[Manage Argos data in Movebank](#)

**What is animal tracking?** Animal tracking data helps us understand how individuals and populations move within local areas, migrate across oceans and continents, and evolve through millennia. This information is being used to address environmental challenges such as climate

## Tracking Data

[Browse tracking data](#)



## The Movebank Data Repository

The **Movebank Data Repository** is a special part of Movebank designed for long-term data archiving. The Repository complements **Movebank's** flexible tools for **sharing, managing, and analyzing** animal movement data throughout all stages of research with a way to formally publish completed research datasets. The data in Movebank are unique observations of animal movements that can never be reproduced—our goal is to make sure these datasets will still be available to answer new questions many decades from now.



Each published dataset, or data package, also has a study in Movebank. If a study in Movebank has been published in the Repository, the DOI and a link will be shown in the **Study Details**.

[Browse the Movebank Data Repository](#)

To read more about the Movebank Data Repository, see the **FAQ**. If you are interested in publishing your own data, contact our data curator Sarah Davidson at [sdavidson@orn.mpg.de](mailto:sdavidson@orn.mpg.de). Published datasets are typically associated with a peer-reviewed journal article. We recommend contacting us after submitting your manuscript but before the proof stage, so we can plan for the data to be published at the same time as the article.

## Submission

[Submit Data Now!](#)

Welcome to Movebank's data repository! Through this repository, Movebank allows users to publish animal tracking datasets that have been uploaded to Movebank ([www.movebank.org](http://www.movebank.org)). Published datasets have gone through a submission and review process, and are typically associated with a written study published in an academic journal. All animal tracking data in this repository are available to the public. We invite you to browse the datasets [here](#).

As of 26.11.2014, the Movebank data repository contains **25 data packages** and **52 data files**.

## Recently Published Data

Spiegel O, Getz WM, Nathan R (2014) Data from: Factors influencing foraging search efficiency: Why do scarce lappet-faced vultures outperform ubiquitous white-backed vultures? (V2). doi:10.5441/001/1.mf903197

van Bommel L, Johnson CN (2014) Data from: Where do livestock guardian dogs go? Movement patterns of free-ranging Maremma sheepdogs. doi:10.5441/001/1.pv048q7v

Spiegel O, Getz WM, Nathan R (2014) Data from: Factors influencing foraging search efficiency: Why do scarce lappet-faced vultures outperform ubiquitous white-backed vultures? doi:10.5441/001/1.pr1vj29n

Griffin L (2014) Data from: Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore [Svalbard data]. doi:10.5441/001/1.5k6b1364

Cabot D (2014) Data from: Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore [Greenland data]. doi:10.5441/001/1.5d3f0664

van der Jeugd H, Oosterbeek K, Ens BJ, Shamoun-Baranes J, Exo K (2014) Data from: Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore [Barents Sea data]. doi:10.5441/001/1.ps244r11

# Research data management at the University of Konstanz – small steps for building up a RDM service

## developing expertise and support

- establishing the position of a research data expert
- Movebank Data Repository: standard service with support by a Movebank data curator
- using the project bwFDM for a deeper analysis of the researchers' needs

## an outlook

- using the organizational structure of an integrated library and computing centre to develop a joint RDM strategy
- stronger cooperation with the research support unit at the university
- assistance in writing RDM plans and reports
- research data management policy for the university (probably)
- institutional data repository (?)
- ... (?)

## 5) Concluding remarks

### political dimension

- e-science- initiative... → recognized importance of RDM in Baden-Wuerttemberg
- but keep in mind: federal strategy versus international cross-linked research

### researcher dimension

- bwFDM bares the existing gap between research and infrastructure
- bwFDM offers a huge and unique dataset of various researcher interests
- certainly not only representative for researchers in Baden-Wuerttemberg

### infrastructure dimension

- several players: computing centres, data centres, libraries, ...
    - a lot of distributed expertise in infrastructure
  - important: cooperation
    - within an institution,
    - and also on a federal/national/international level
- but keep in mind: this is not very well established at the moment

# Concluding remarks

## overall estimation

- it will be a challenge to establish a researcher-driven effective infrastructure fulfilling all the various needs
- the bwFDM survey could – hopefully - form the nucleus for a broad RDM infrastructure deployment in Baden-Wuerttemberg
- we should start on several levels: local (example Konstanz), federal, international
- a nice example: TROLLing: Open Data for Linguistics  
our Konstanz linguistic professor (running also an OA-journal) delightedly informed our library director last week about the promoting youtube video, happy to use this discipline specific data repository in future

# Sources

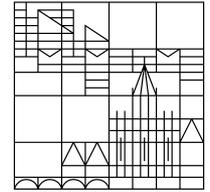
**e-science-initiative in Baden-Wuerttemberg:** [http://mwk.baden-wuerttemberg.de/uploads/media/066\\_PM\\_Anlage\\_E-Science\\_Web.pdf](http://mwk.baden-wuerttemberg.de/uploads/media/066_PM_Anlage_E-Science_Web.pdf)

**bw-projects:** <http://www.alwr-bw.de/kooperationen/>

**bwFDM:** <http://bwfdm.scc.kit.edu/>

**Movebank Data Repository:** <https://www.datarepository.movebank.org/>

Universität  
Konstanz



**Thank you  
very much!**

**Karlheinz Pappenberger**

University of Konstanz – KIM

subject librarian for economics and statistics  
and specialist for research data management

[karlheinz.pappenberger@uni-konstanz.de](mailto:karlheinz.pappenberger@uni-konstanz.de)