RDMG – Research Data Management Group at Freiburg University

Round Table RDM, Academy of Sciences in Heidelberg

13/07/2018

Dirk von Suchodoletz, eScience dept. of the Computer Center

Albert-Ludwigs-Universität Freiburg

RDMG

Research Data Management Group

UNI FREIBURG

Institutional RDM

- Rising demand for Research Data Management
  - Science got mainly digital
  - Creation, production of data involves significant costs
  - Reuse of data
  - End-to-end data management concepts
- Requirements set by funding institutions

  → RDM becomes new column in research organisations framework beside library, arts collections, zoology exhibition, ...

FAIR principles

- Requirements towards research data defined in 2105 by major institutions

  - Findable

  - Accessible

  - Interoperable

  - Reusable

- Idea to support knowledge discovery and innovation, support data and knowledge integration, promote sharing and reuse of data

- Discipline independent and allow for differences in disciplines

RDMG

Funding requirements

- Requirements will get more and more implemented on national level
  - Universities should have a data management policy
  - DFG started to require Data Management Plans for a while (as well as EU Horizon 2020 is requiring it already)
  - Same in the ongoing Excellency Strategy
  - In the future: New research projects might get formally rejected if no (proper) data management got described and implemented
  - Data management will become part of project evaluation in the future

RDMG

Funding requirements

- Increasingly formal requirements to get research funded
  - e.g. COMMISSION RECOMMENDATION (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information:

```
1. Member States should set and implement clear policies (as detailed in
national action plans) for the dissemination of and open access to
scientific publications resulting from publicly funded research. Those
policies and action plans should provide for:
— concrete objectives and indicators to measure progress,
— implementation plans, including the allocation of responsibilities and
appropriate licensing,
— associated financial planning.

Member States should ensure, in compliance with the EU acquis on
copyright and related rights, that as a result of these policies or
action plans: ...
```

Creation of the RDMG

- University of Freiburg quite late in the process but various groups were active already before

  - Research Data Management Group (RDMG) - Cooperation of the university library, the Freiburg Research Services and the computer center

- Computer center as core service provider for research infrastructures

  - Storage for various purposes

  - Tape storage for backup and disaster recovery

  - Facilities to host large scale IT equipment

  - Focus on technical expertise but less experience with data curation, annotation, …

Linearization of the Lifecycle
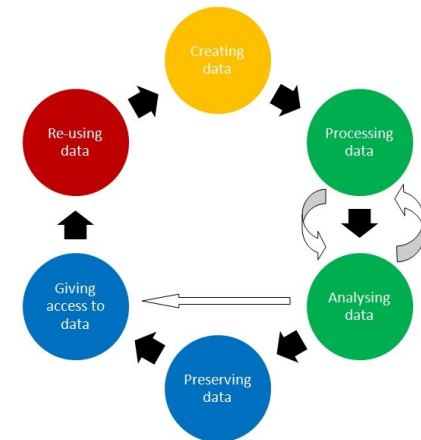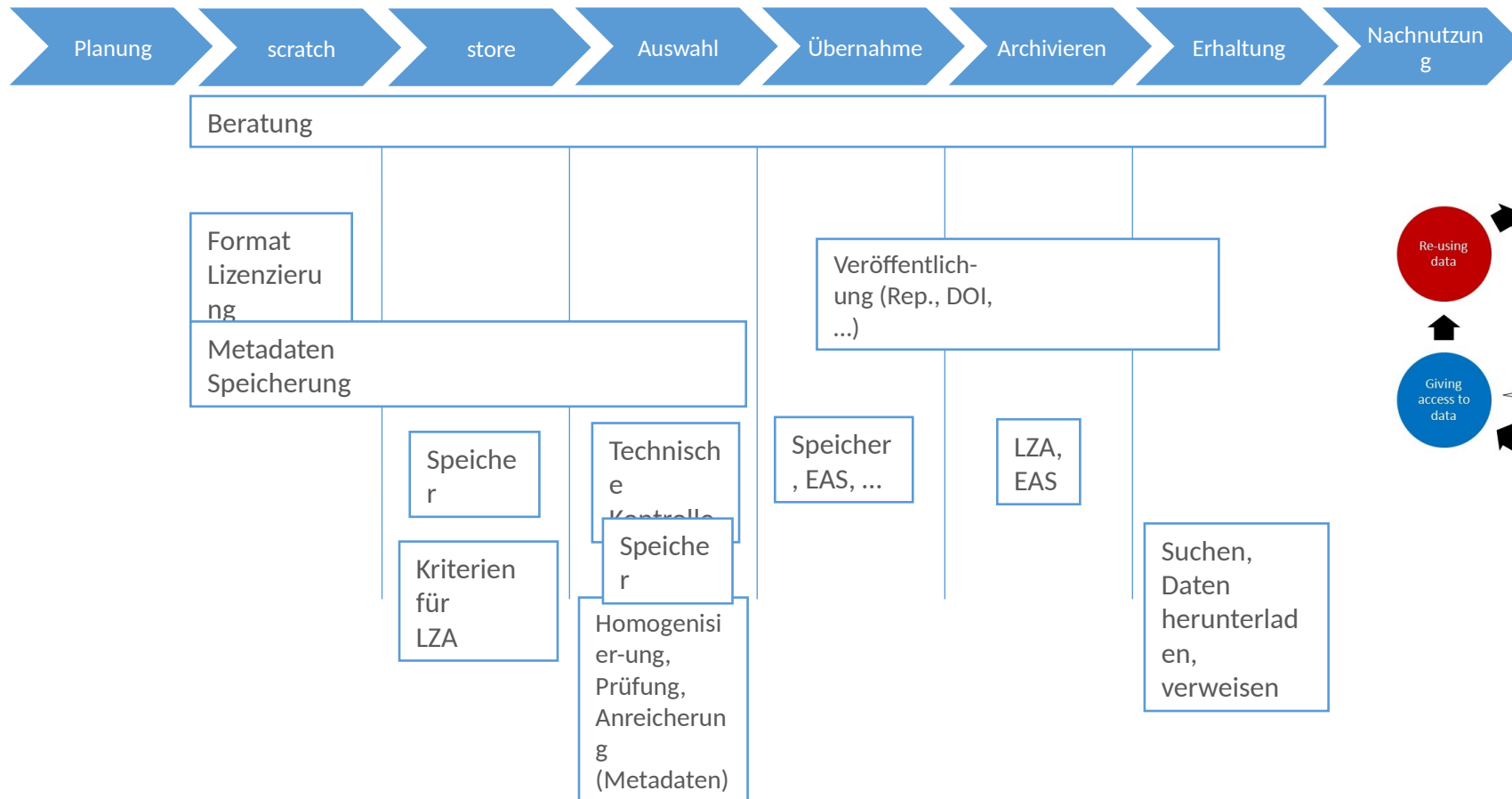
- **Tasks attachable to phases in the lifecycle**

RDMG: Communication & material

- Development of a "RDM strategy for the university" by the eScience groups of the university library and the computer center (draft made available)

- OTRS queue for requests and consultation: *fdm@mail.uni-freiburg.de*

- Internal e-Learning platform for "Weiterbildung", see https://wb-ilias.uni-freiburg.de/goto.php?target=frm_130918_10125&client_id=unifreiburgwb

- State-wide: https://forschungsdaten.info and the RDM WG (AK FDM)

Research Infrastructures

Research Infrastructures

- Computer Center operates significant infrastructures
  - VMware ESX Virtualization
  - General data storage for homes & groups
  - bwCloud IaaS
  - NEMO HPC
  - TSM Tape Disaster Recovery
- Future: bwSFS
  - Storage for Science as system for the introduction of university level research data management system

From NFS to RDM system

- Traditional file servers not enough to cope with modern requirements
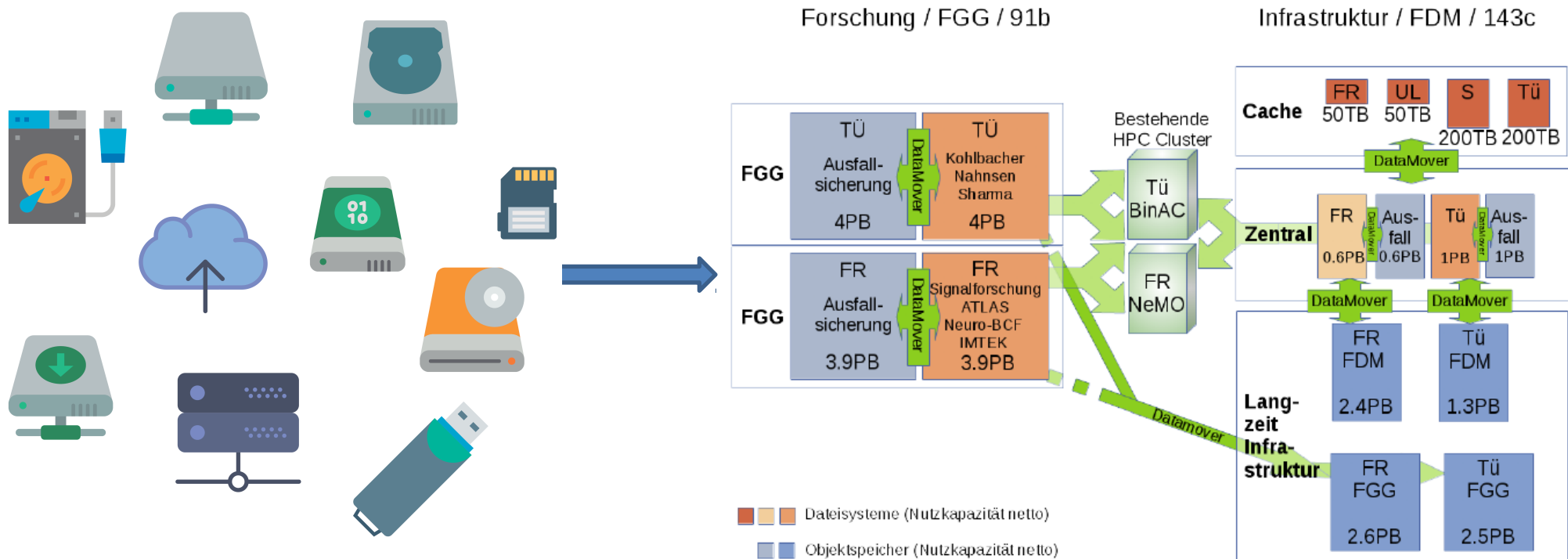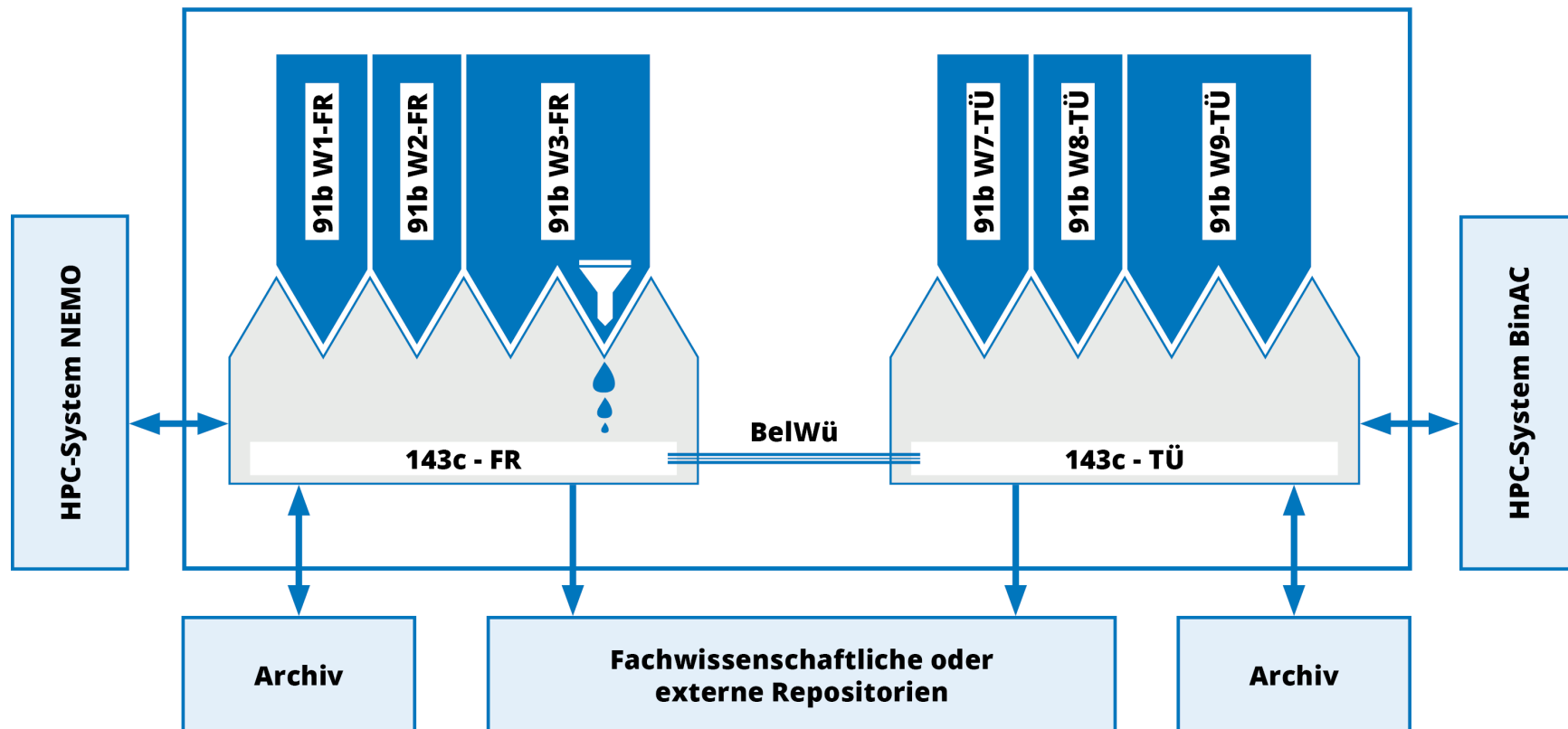
Envisioned core RDM services

- **Repository services like e.g. iRODS**
  - Evaluated within the bwDATAbib project
  - Provide the linking of data to it's meta data
- **Object store or filesystem services**
- **Versioning services like e.g. GIT**
- **(Longterm) Archival services**

RDMG

Initial Financing

- **Structure of the DFG grant**
  - General part covered by 143c

Meta Data & Decision Making

Technical and Scientific Meta Data

- Required for both automated and manual decision making

- Meta Data required for the Data Mover (some information only relevant for technical purposes)

- Meta Data required for re-use of data sets (defined by respective scientific community)

- Meta Data required to identify research projects (to link to e.g. grants in Research Information Systems)

RDMG

Definition of Services & Meta Data

- *Scientific communities* define their RDM / storage needs
  - Amount of data and expected time spans
  - Amount of data over the data life cycle
  - Access methods to the stored data
  - Required scientific Meta Data
  - Policies for access and re-use

RDMG

Data Cite and Object Access

- Data will become more and more part of scholarly communication and publication

- DataCite – using DOIs (Digital Object Identifiers or handles for less static objects)

- DOIs should be persistent and reference data set independent of its actual location within the SfS system

- Open Archives Initiative (OAI) Protocol for Metadata Harvesting

- Many libraries use OAI to exchange information between repositories

RDMG

Governance & Financing

Governance

- Very diverse requirements by different groups
- *University library & computer* center can not decide on data lifetime, validity, worth, ...
  - Costs associated with amount of data (single data sets, storage capacity required by a project, faculty, ...)
  - Keeping costs under control requires regular optimizations and cleanup
  - Cost models need to be created in the mid and long run

RDMG

Financing Model

- Core institutions of a university usually have their own budget

- Often research & service launch funds available from state or science funding institutions

- Not sustainable in the long run
  - Costs need to be covered by the university (and might get re-financed by instituts, projects, ...)

RDMG

Re-Financing

- **Possible for researchers to get grant money for data management**
  - Some funding institutions require open data and expect data to be published

- **Challenges**
  - Data ownership changing over time
  - Access policy might change over time (e.g. from restricted to open)

- **Faculties need to get involved e.g. via academic self governing structures**
  - Create committees to decide on RDM related issues

RDMG

Sum-up / Conclusion

- **RDM becoming increasingly relevant**
  - Future assets of a university
- **Challenges**
  - Various preexisting solutions need to get properly included, referenced, handled
  - (Internationally) Cooperating scientists vs. local institutions
  - Very diverse requirements
  - ...

  !Thank you very much for your attention!