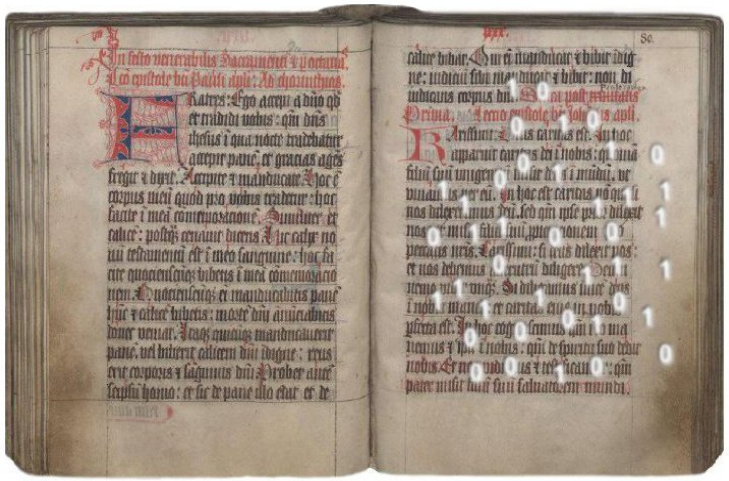
DARIAH-DE

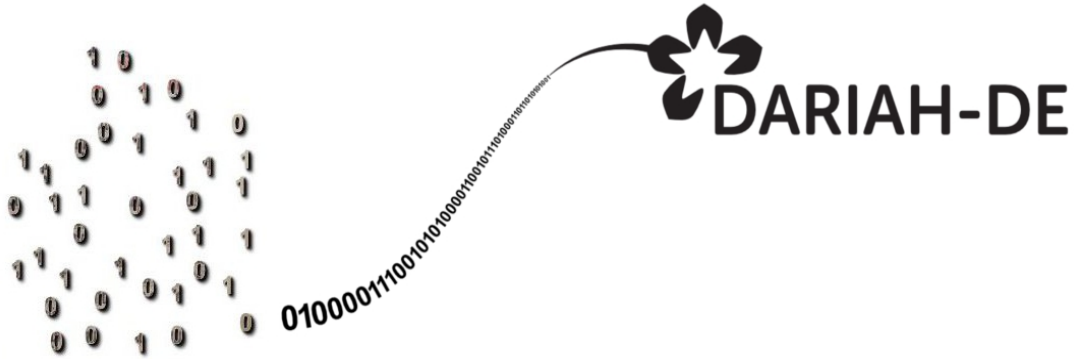
Digital Research Infrastructure
for the Arts and Humanities

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

The DARIAH Bit Preservation Service for Humanities Research Data





de.dariah.eu

DARIAH In A Nutshell

- ESFRI (**E**uropean **S**trategy **F**orum on **R**esearch **I**nfrastructures) project for the arts and humanities
- Based on national contributions

DARIAH-DE
Digital Research Infrastructure for the Arts and Humanities

- Scope:
  - Build-up of a **sustainable research infrastructure** in close, international cooperation
  - Enhancement of **digital research methods**
  - **Access** to **digital research data** in Europe
  - **Support** of scholars

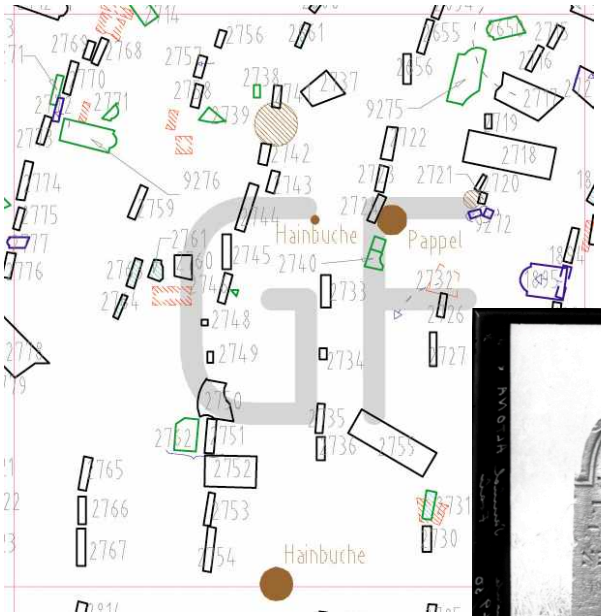
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

- Participants: archaeology, musicology, Jewish Studies, history, philology, art history, philosophy, literary studies, ...
  + computer science + data centers

Examples Of Research Endeavors

Digitale Edition - Jüdischer Friedhof Hamburg-Altona, Königstraße (1621-1871 / 5988 Einträge): Inv.-Nr. 3361
URL: http://www.steinheim-institut.de/cgi-bin/epidat?function=Ins&sel=hha&inv=3361 (2013-02-21)

פה נקבר	*Hier ist begraben*
הישיש כמר	*der Hochbetagte, der geehrte Herr*
שמואל בר יהו	*Schmuel, Sohn des Jehuda,*
דא שעיי נתישבה	*von dem die heilige Gemeinde*
ק"ק אלטנא נפטר	*Altona gegründet wurde, verschieden*
יו' ג' ר"ח אלול שפ"א	*Tag 3, Neumond Elul 381*
לפ"ק תנצב"ה	*der kleinen Zählung. Seine Seele sei eingebunden in das Bündel des Lebens*



Examples Of Research Endeavors

Public Library and Archive of Trier
Ms 1108/55 4° 6v and 7r

New research opportunities and questions by

- access to research data
- processing and analyzing a huge amount of research data

→ Projects rely on accessible, long-term data storage

→ Heterogeneous data and results need to be preserved for the future

Bit Preservation And Long-term Archiving

Data Curation - Interpretability

- Creation
- Object management

Content Preservation - Readability

- Versioning + provenance
- Data formats

Bit Preservation

- Integrity preservation (checks, error correction codes)
- Replication

Bit Preservation And Long-term Archiving

Data Curation - Interpretability

- Creation
- Object management

Content Preservation - Readability

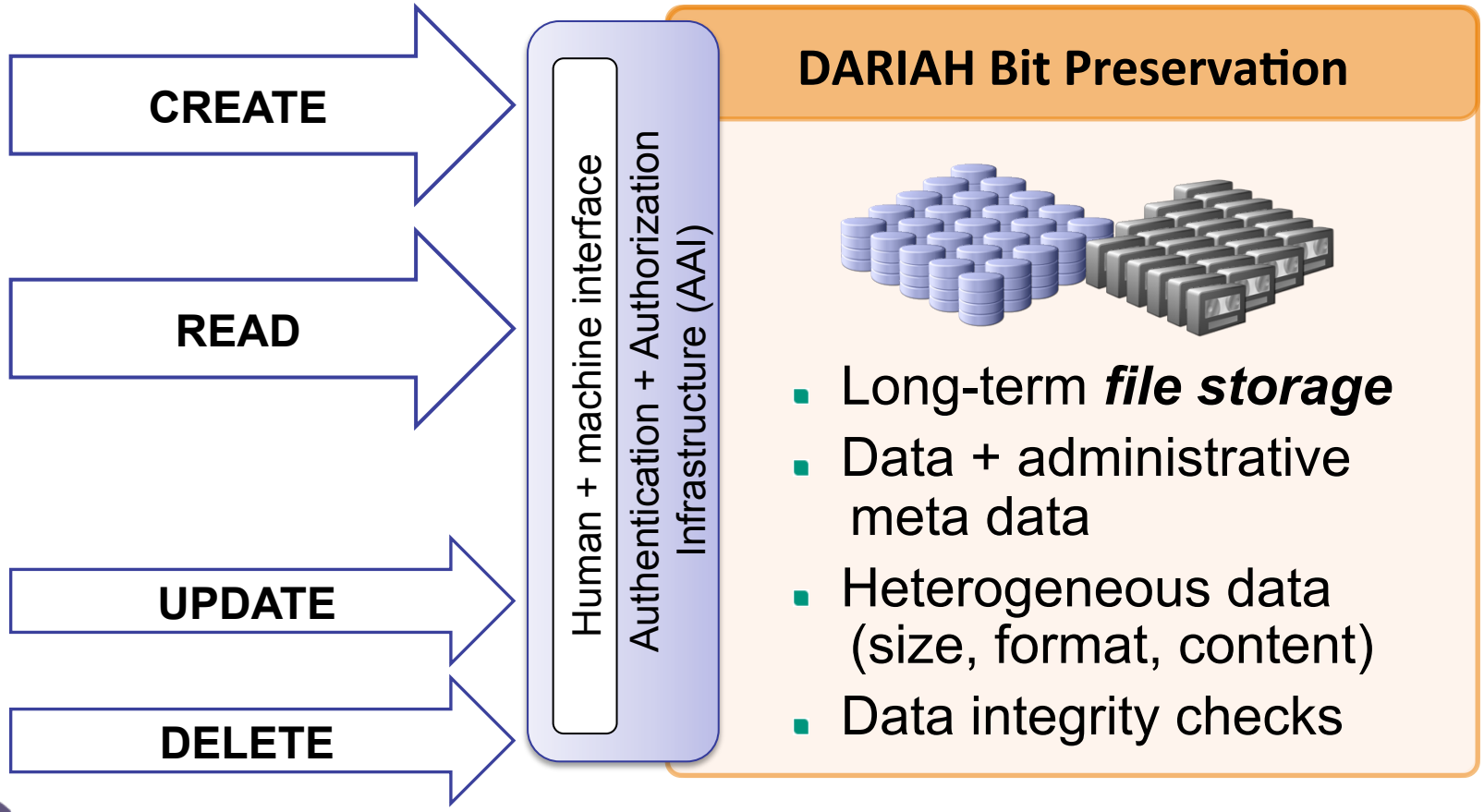
- Versioning + provenance
- Data formats

Bit Preservation

- Integrity preservation (checks, error correction codes)
- Replication

Features Of The DARIAH Bit Preservation

The DARIAH Bit Preservation aims to design and implement a system for a sustainable, safe and persistent storage of research data.



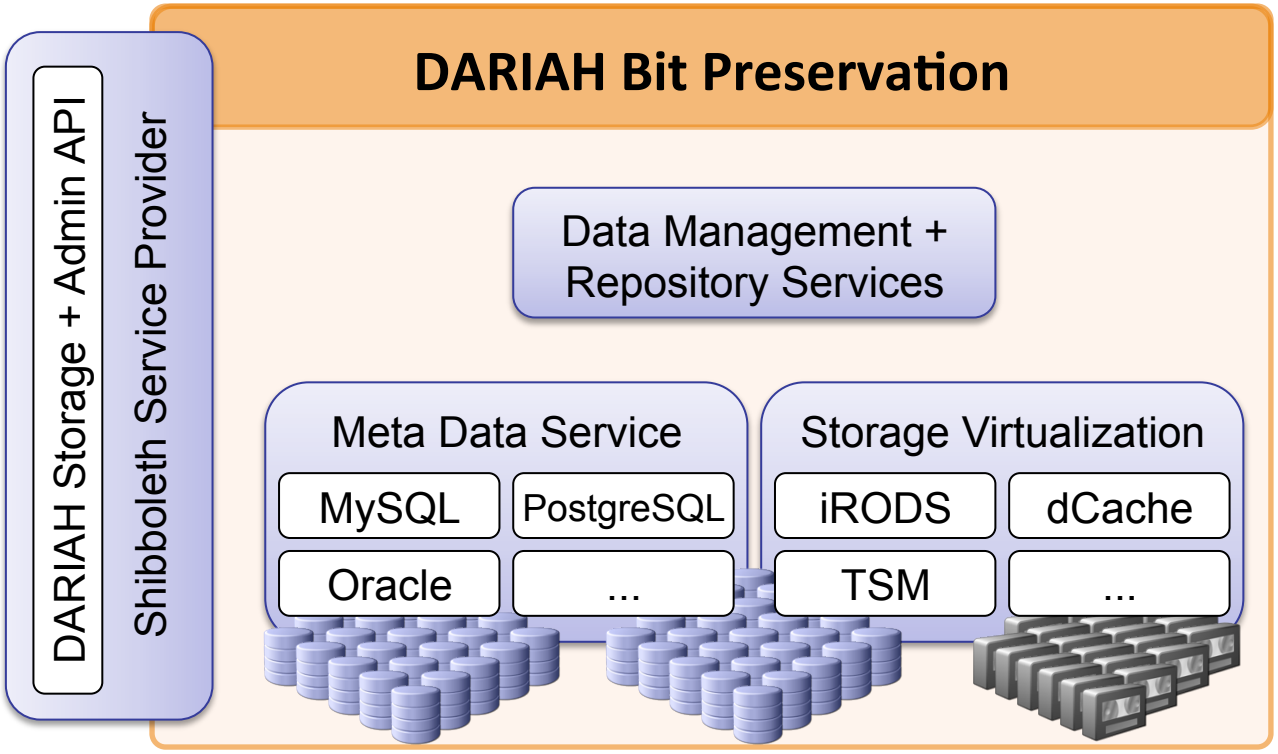
Standard-based Interfaces

Requirements:

- As easy-to-use as possible
- Hide the complexity of the DARIAH Bit Preservation
- Basic functionalities for file handling
- Basic configuration of Bit Preservation mechanisms

DARIAH Storage API (application programming interface)	**Bit Preservation Admin API (application programming interface)**
RESTful + HTTP-based	RESTful + HTTP-based
Storage functionalities	Bit Preservation functionalities
	Administrative interface
	Information about data (e.g. number of replicas, last integrity check)
	Configuration of Bit Preservation

Architecture



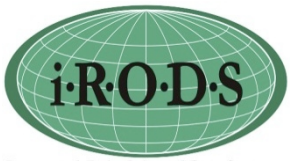
- Separation of functionalities
- Independent from storage backend
- Standardized interface

Sustainability + Usability

Existing Storage Backend Adapters

- Storage provided by KIT, GWDG, RZG, FZJ
- KIT: 50 TB in Large Scale Data Facility (SCC, CN)

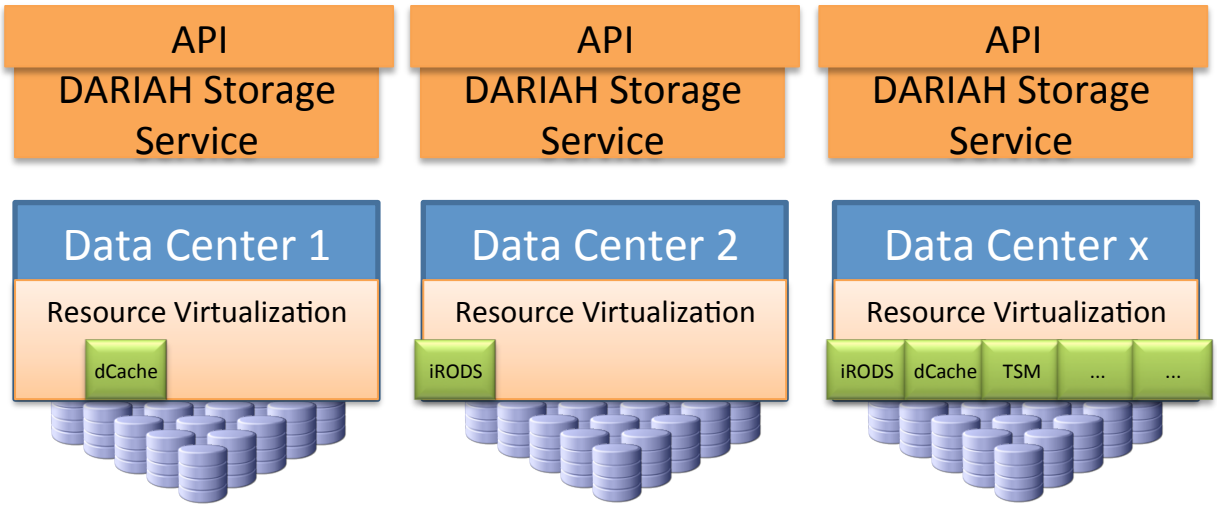
- ***iRODS***: integrated Rule-Oriented Data System
- Open-source data management system
- Logical namespace independent from storage resources
- Rules to implement workflows / policies



- ***Hard drive***

- ***dCache*** (thanks to Xavier and Doris)
- Caching system for frequently requested data
- Enables convenient access to a tape system

Research + Development Challenges

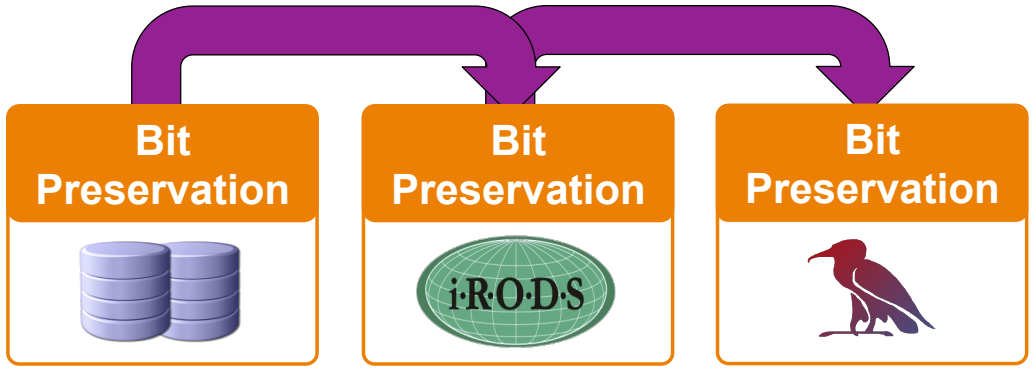


- Definition + implementation of a storage service independent of the underlying storage backend
- Proof of interoperability of the storage backend

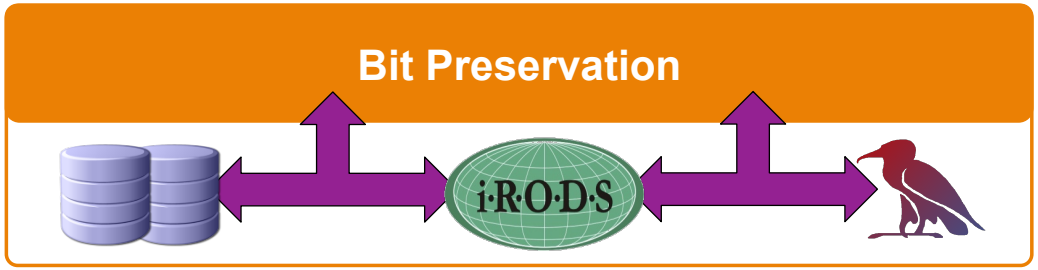
Currently Under Evaluation

- ***Data migration*** – use case „Virtual Scriptorium“ with ~ 5 TB
- Implementation of migration tools and evaluation (stability, performance, impact on Bit Preservation mechanisms)

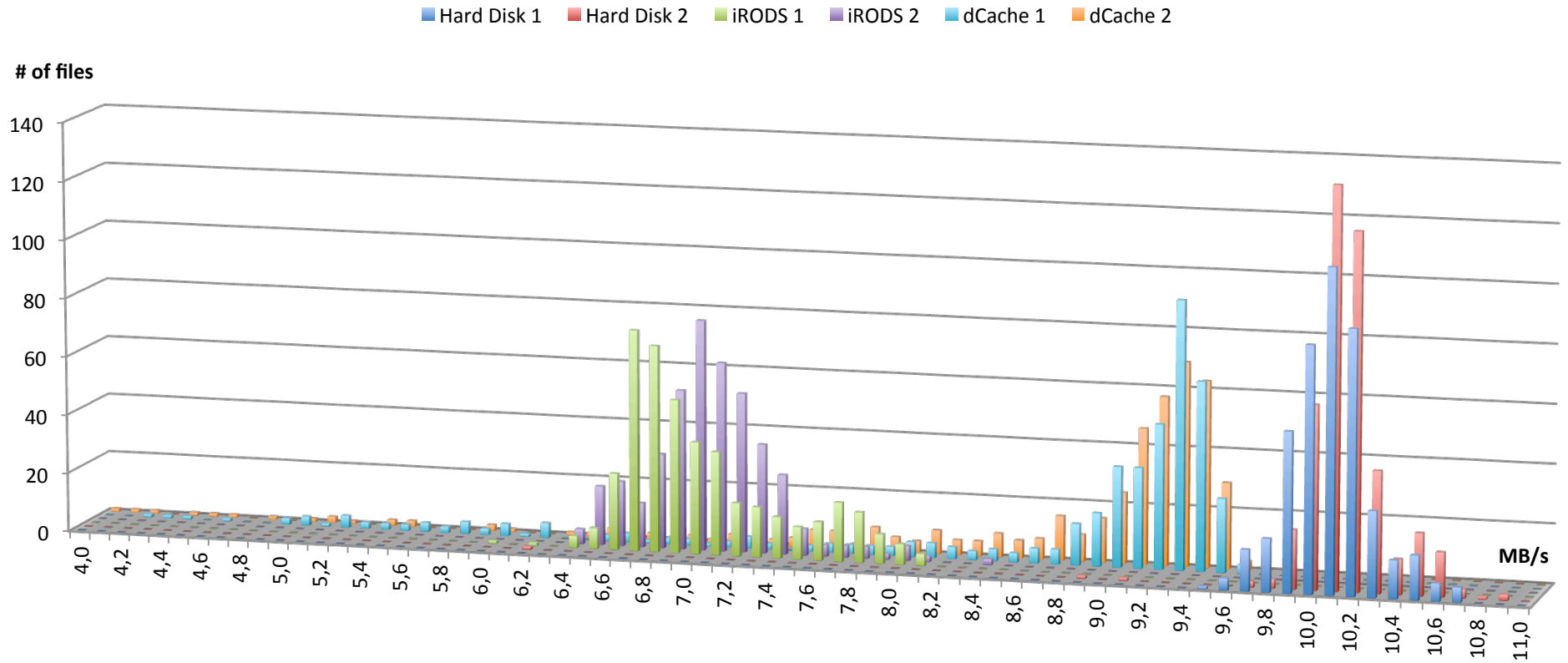
- ***Approach 1:***
  migration outside the Bit Preservation



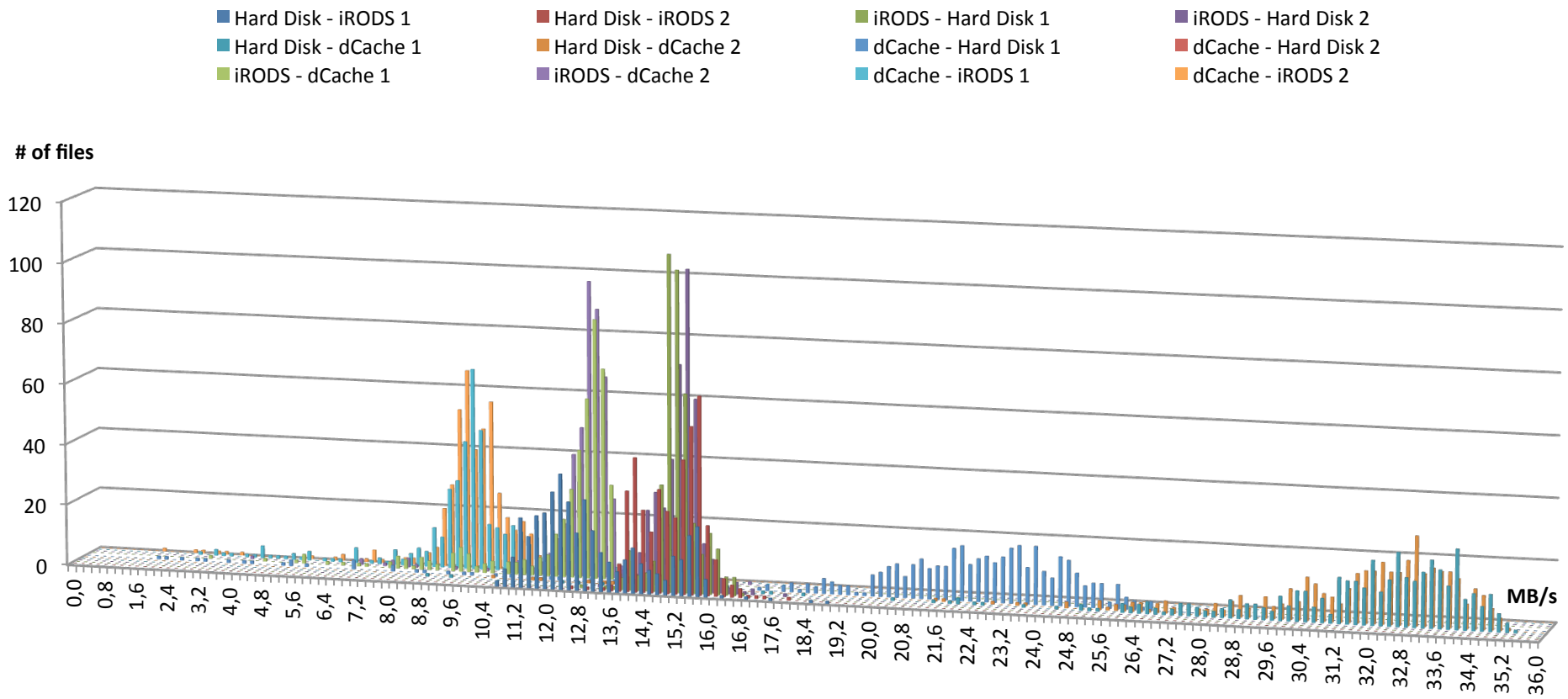
- ***Approach 2:***
  migration inside the Bit Preservation



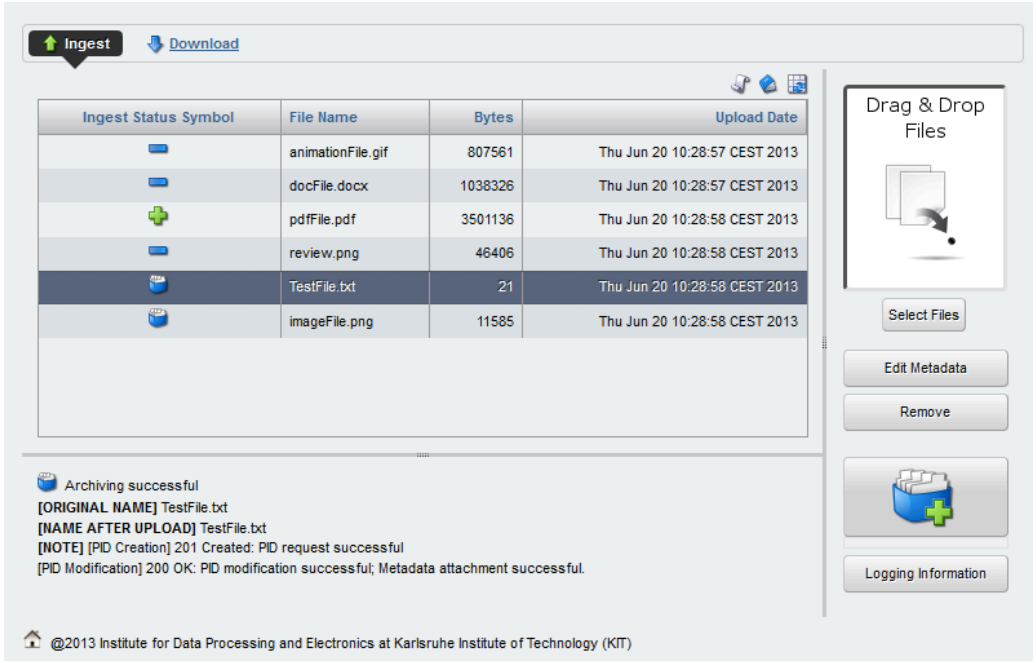
Migration Outside The Bit Preservation



Migration Inside The Bit Preservation



dawa – data web application



- Service for long-term, referenceable file storage
- Combination of storage, authorization, authentication and PID service
- ***Directly usable*** for the humanities scholar
  - Automatic workflows
  - Intuitive, graphical user interface

Bit Preservation Facts

1. Standard-based interface with ***consent of six institutions*** (including four data centers) specified

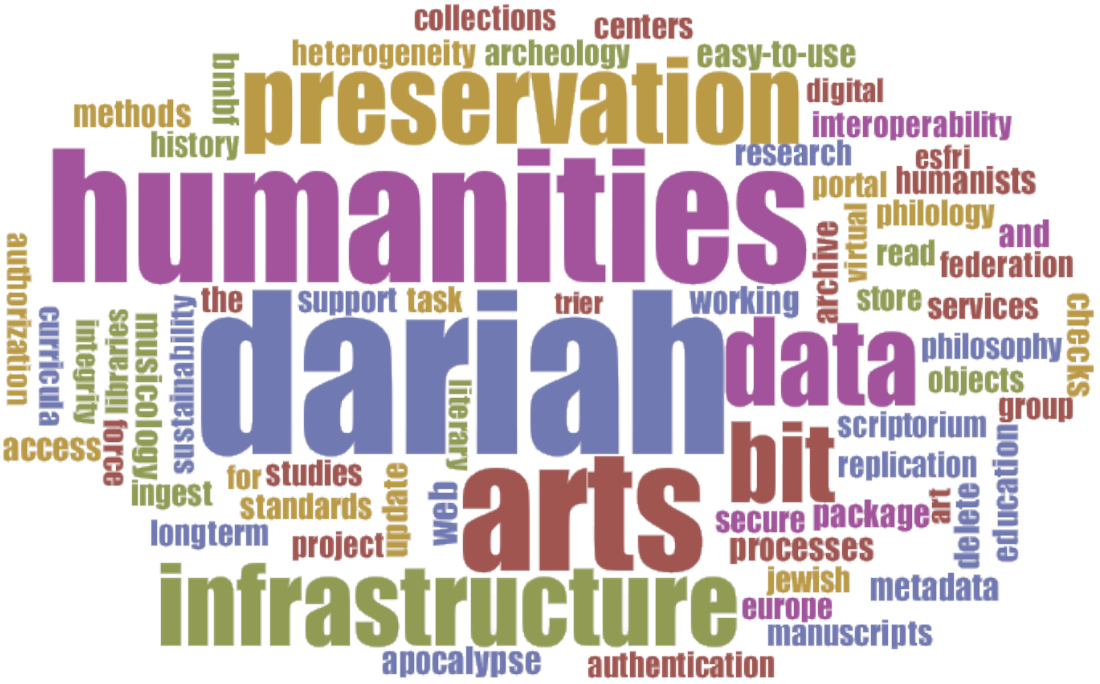
2. Used and ***trusted by higher-level services*** (e.g. GeoBrowser, AiB, dawa)

3. Implementation running ...
   ... since 2012 at KIT
   ... since 2013 at GWDG, FZJ, RZG

4. Although designed for small data successfully handled ~ 80 TB

5. Transferable to and ***reusable by other communities*** (e.g. libraries by dawa‘s automatic assignment of PIDs)

Starting A Phd



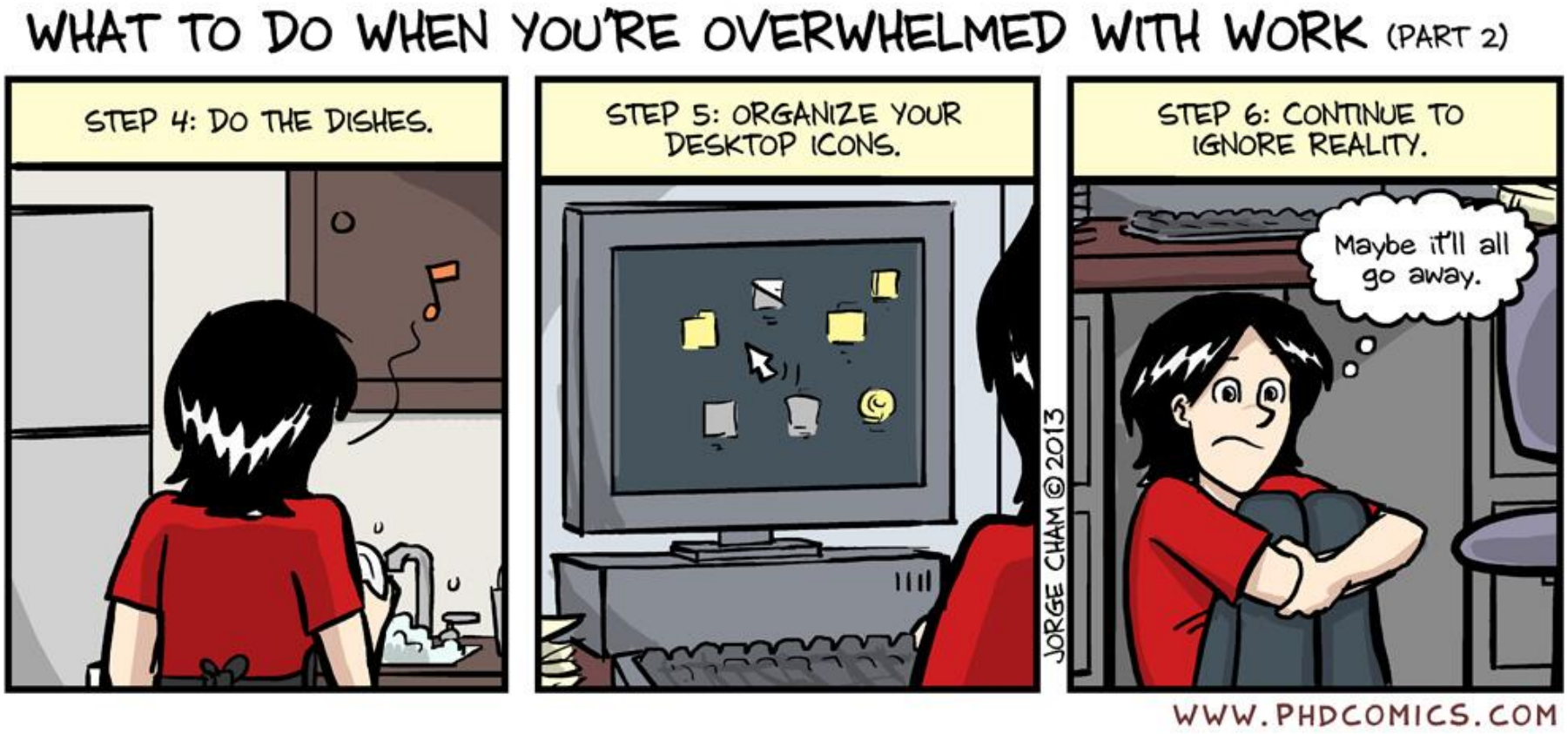
DARIAH-DE

DARIAHner At Work

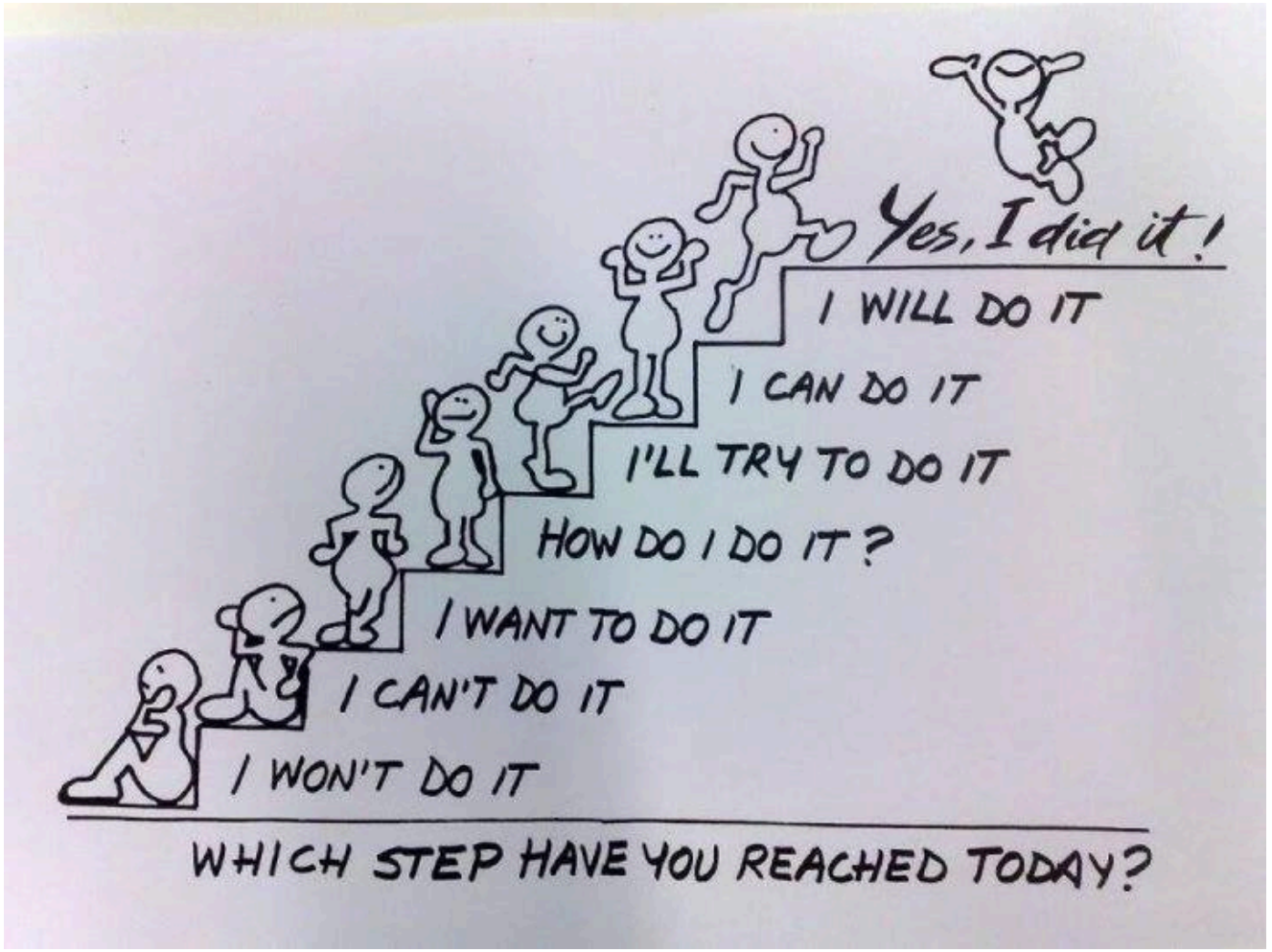
Lots Of Stuff To Do

DARIAH-DE

Lots Of Stuff To Do



What Helped At Least A Little ;-)



DARIAH-DE