

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



GEFÖRDERT VOM



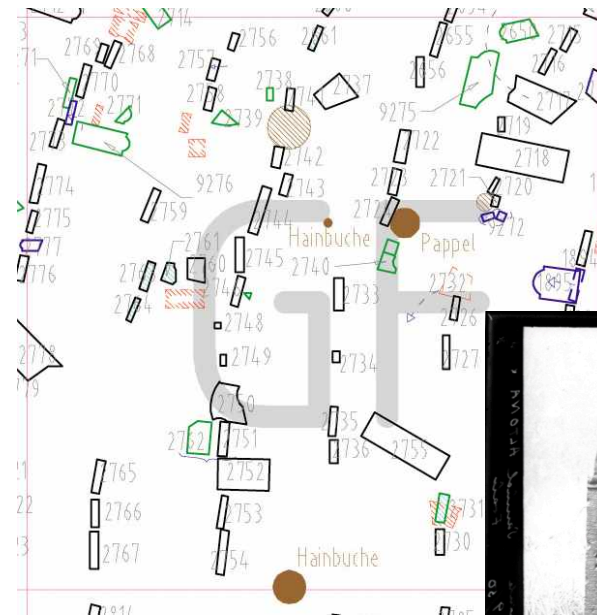
Bundesministerium
für Bildung
und Forschung

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



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

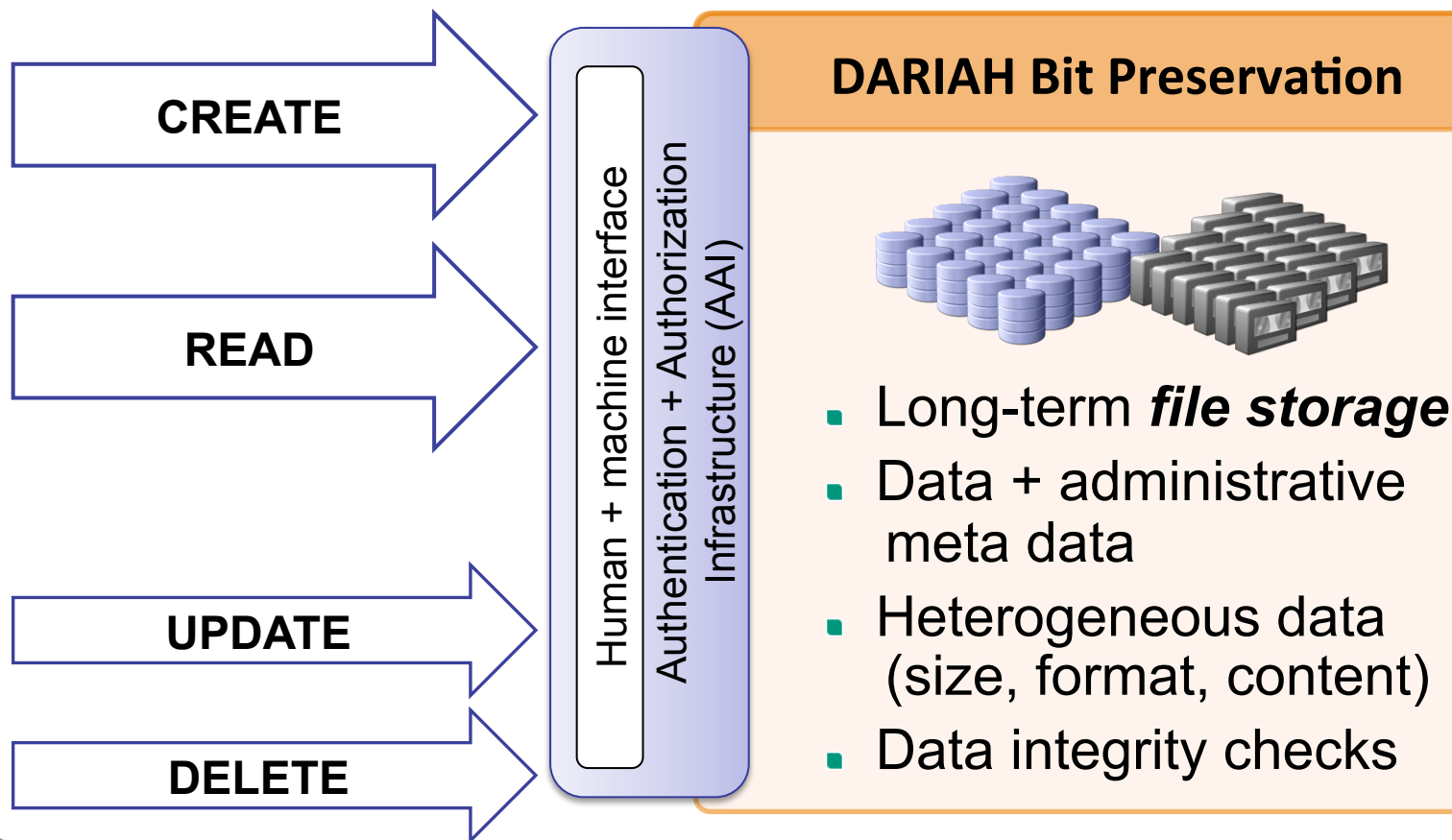
- Versioning + provenance
- Data formats

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.



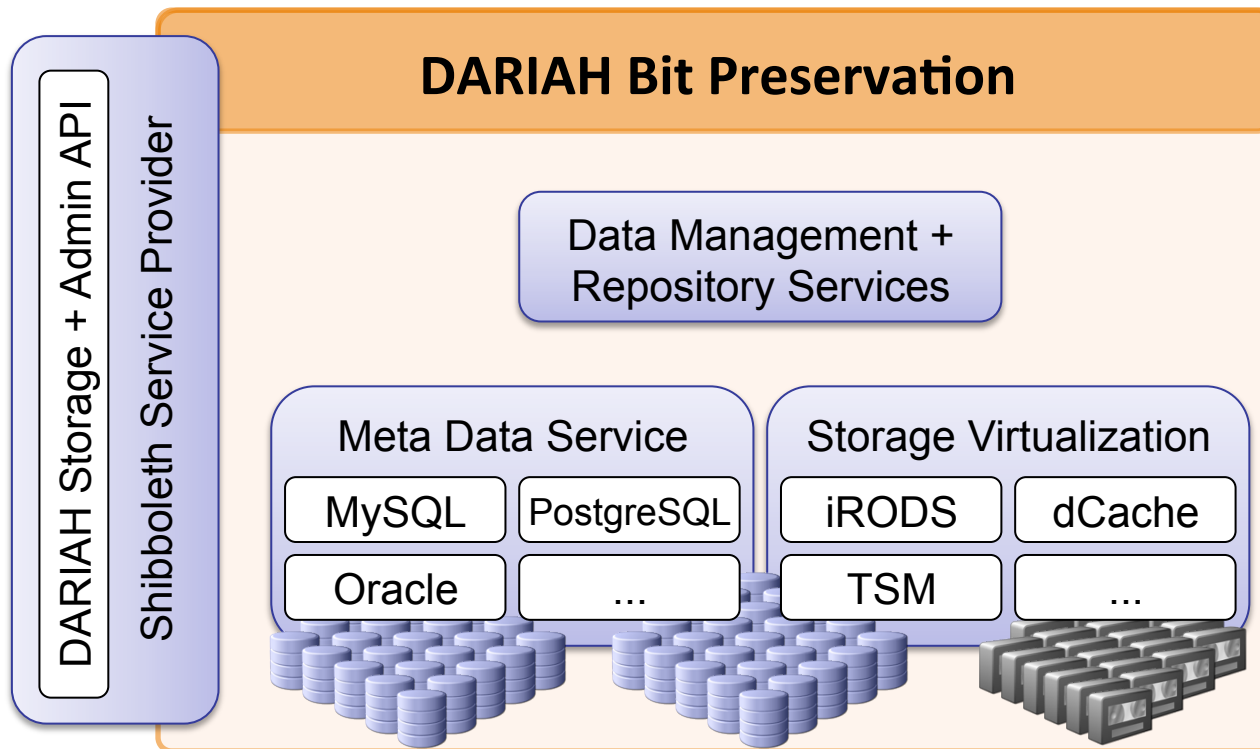
Standard-based Interfaces

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

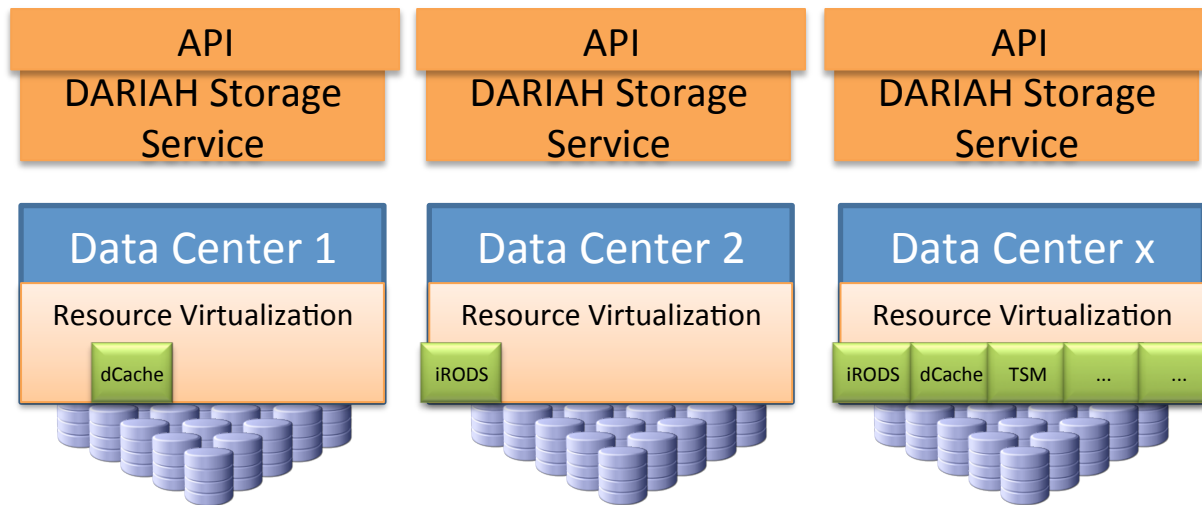
Sustainability + Usability

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

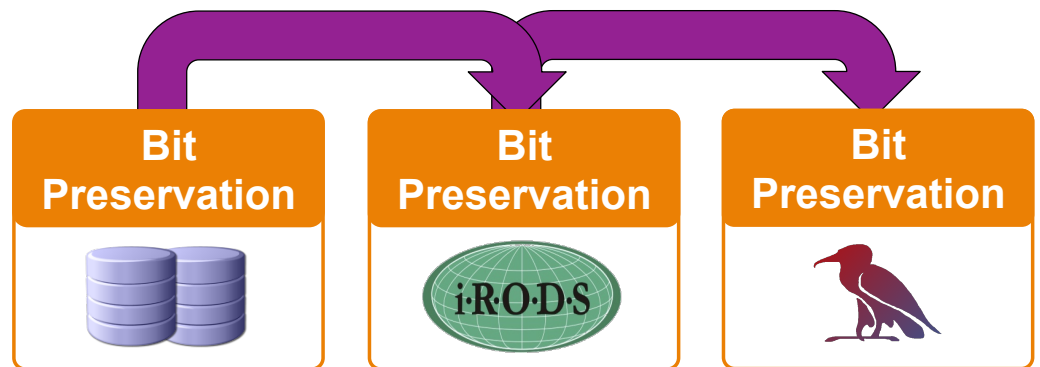


- Definition + implementation of a storage service independent of the underlying storage backend
- Proof of interoperability of the storage backend

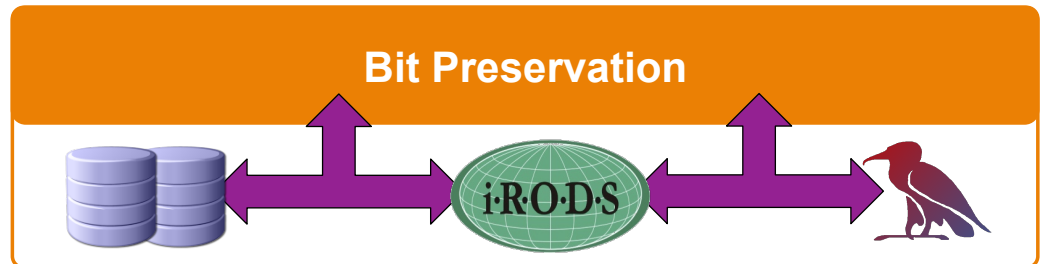
Currently Under Evaluation

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

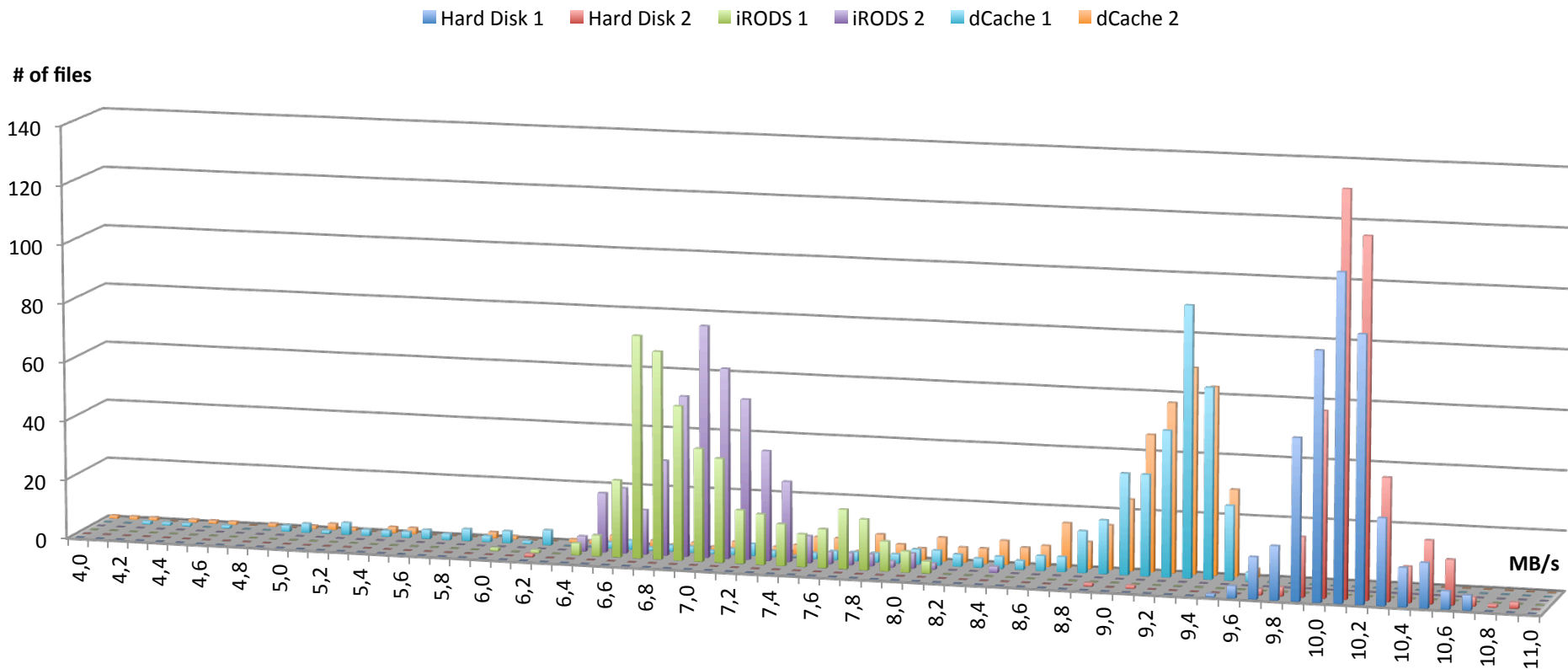
- **Approach 1:**
migration outside
the Bit Preservation



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

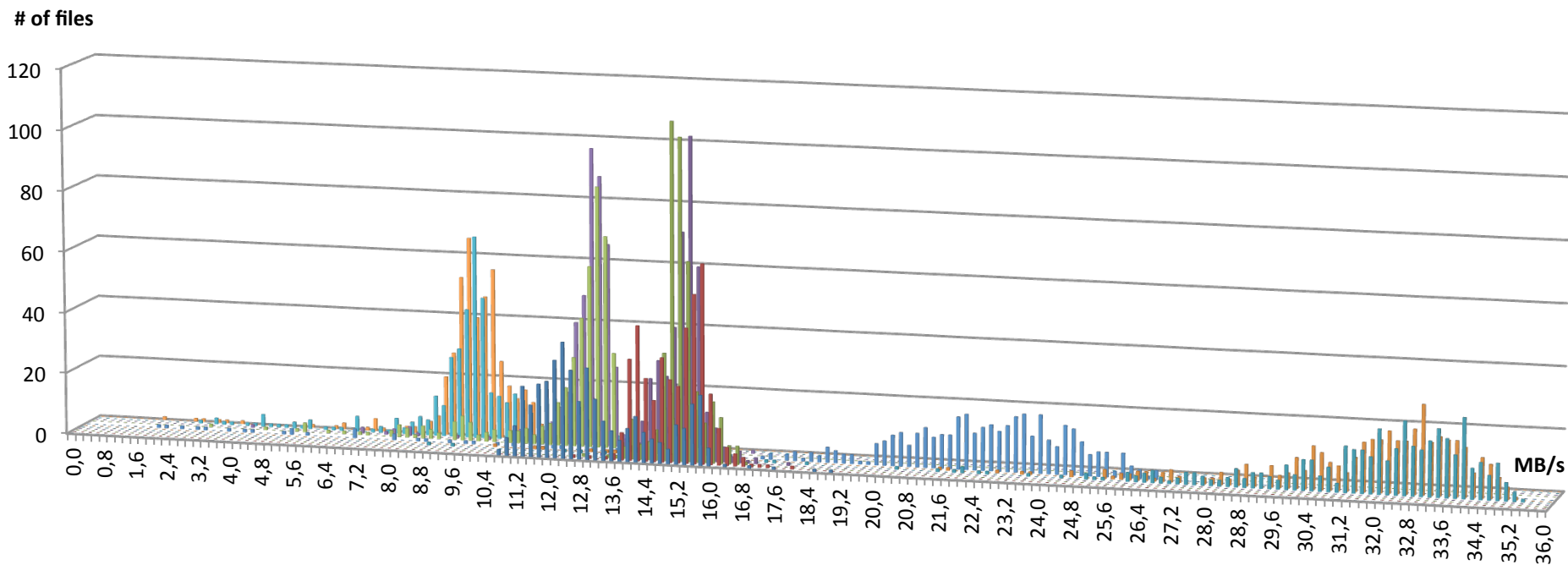


Migration Outside The Bit Preservation

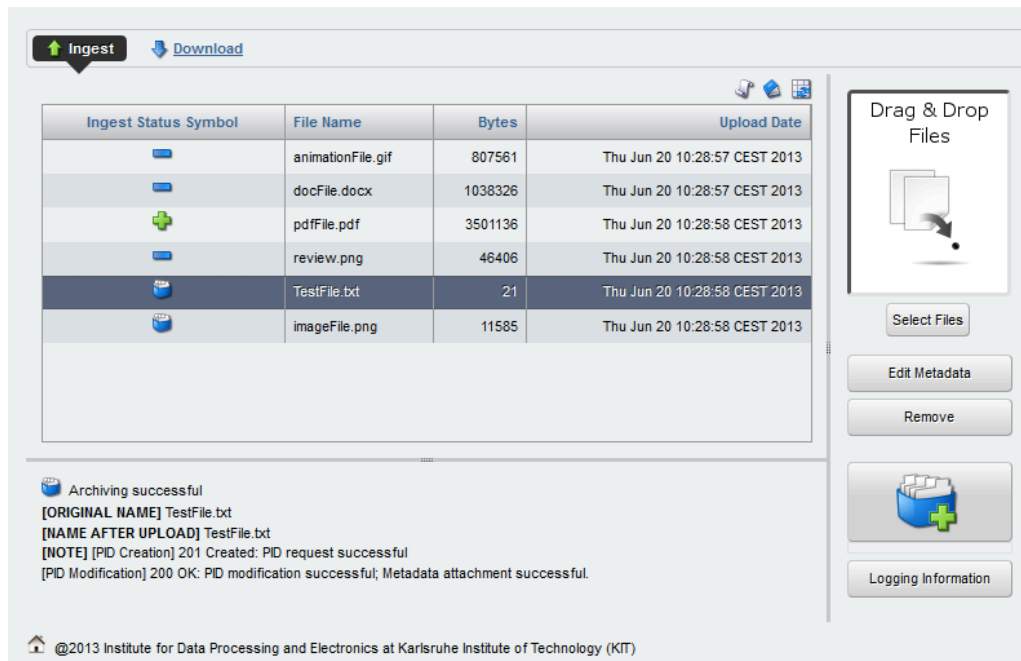


Migration Inside The Bit Preservation

- Hard Disk - iRODS 1
- Hard Disk - dCache 1
- iRODS - dCache 1
- Hard Disk - iRODS 2
- Hard Disk - dCache 2
- iRODS - dCache 2
- iRODS - Hard Disk 1
- dCache - Hard Disk 1
- dCache - iRODS 1
- iRODS - Hard Disk 2
- dCache - Hard Disk 2
- dCache - iRODS 2



dawa – data web application



The screenshot displays the DAWA web application interface. At the top, there are 'Ingest' and 'Download' buttons. Below is a table with columns for 'Ingest Status Symbol', 'File Name', 'Bytes', and 'Upload Date'. The table contains several rows of file uploads, with 'TestFile.txt' highlighted. To the right of the table is a 'Drag & Drop Files' area with a 'Select Files' button and other options like 'Edit Metadata', 'Remove', and 'Logging Information'. Below the table, there is a message indicating successful archiving and PID creation/modification.

Ingest Status Symbol	File Name	Bytes	Upload Date
	animationFile.gif	807561	Thu Jun 20 10:28:57 CEST 2013
	docFile.docx	1038326	Thu Jun 20 10:28:57 CEST 2013
	pdfFile.pdf	3501136	Thu Jun 20 10:28:58 CEST 2013
	review.png	46406	Thu Jun 20 10:28:58 CEST 2013
	TestFile.txt	21	Thu Jun 20 10:28:58 CEST 2013
	imageFile.png	11585	Thu Jun 20 10:28:58 CEST 2013

Archiving successful
[ORIGINAL NAME] TestFile.txt
[NAME AFTER UPLOAD] TestFile.txt
[NOTE] [PID Creation] 201 Created: PID request successful
[PID Modification] 200 OK: PID modification successful; Metadata attachment successful.

@2013 Institute for Data Processing and Electronics at Karlsruhe Institute of Technology (KIT)



- 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)

Implementation running ...

3. ... 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

SCIENTIST IN TRAINING

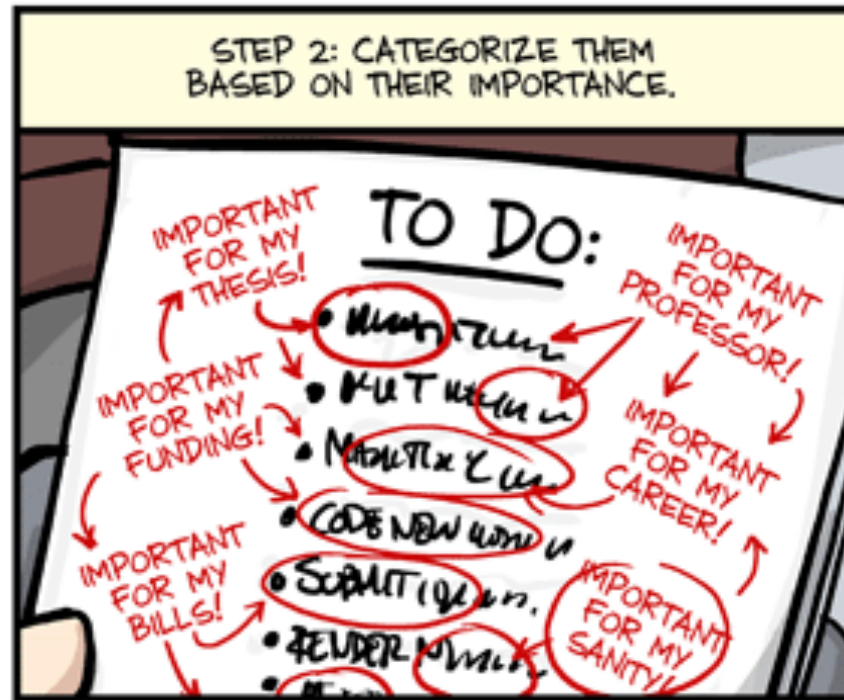


DARIAHner At Work



Lots Of Stuff To Do

WHAT TO DO WHEN YOU'RE OVERWHELMED WITH WORK



JORGE CHAM © 2013

WWW.PHDCOMICS.COM

Lots Of Stuff To Do

WHAT TO DO WHEN YOU'RE OVERWHELMED WITH WORK (PART 2)



JORGE CHAM © 2013

WWW.PHDCOMICS.COM

What Helped At Least A Little ;-)

