DLCL Key Technologies: October 2015

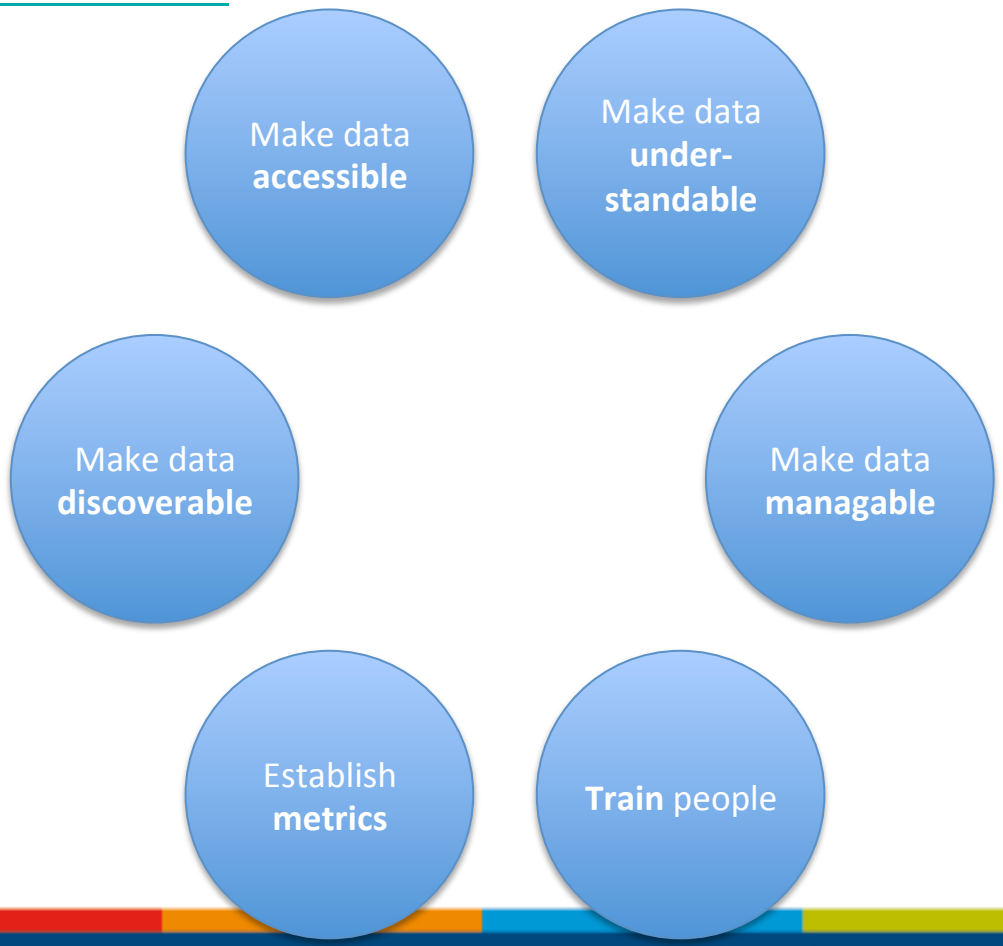
Rainer Stotzka, Swati Chandna, Richard Grunzke, Volker Hartmann, Michael Hausmann, Jürgen Hesser, Thomas Jejkal, Ralph Müller-Pfefferkorn, Michelle Pfeiffer, Francesca Rindone, Danah Tonne, Xiaoli Yang, Sasa Vondrous, Eberhard Schmitt, Margund Bach, Ajinkya Prabhune, Armin Volkmann, Hjalte Raun, Kevin Geggus, Anil Keshav, Aaron Zweig, Hasebullah Ansari

LSDMA

G8 Principles

White Paper: 5 Principles for an Open Data Infrastructure
https://epubs.stfc.ac.uk/work/12236702

- searchable → create useful metadata
- accessible → deposit in trusted repository and use PIDs
- interpretable → create metadata, register schema, and semantics
- re-usable → provide contextual metadata
- persistent → provide persistent repositories

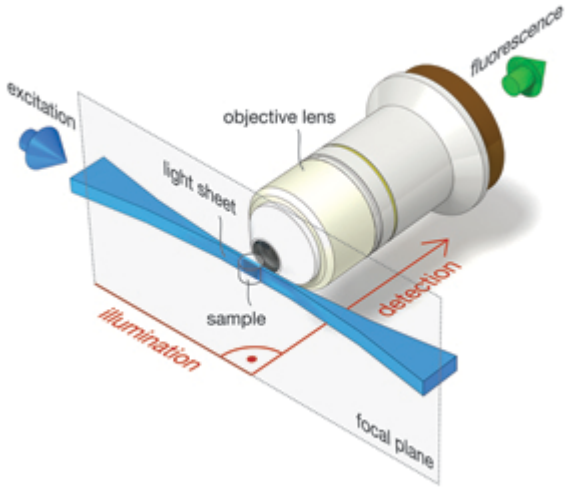


Subprojects

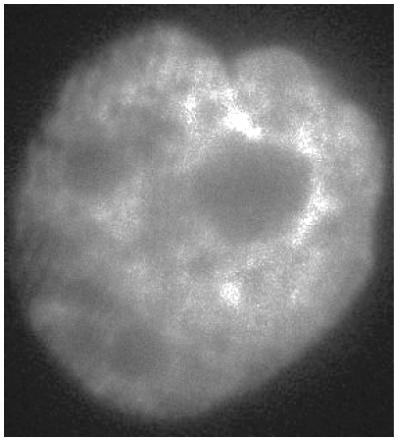
- ***Light Optical Nanoscopy*** (Heidelberg, Mannheim, Mainz)
- ***High Throughput Microscopy***:
  - Selective Plane Illumination Microscope (Karlsruhe)
  - Gen Scans (Dresden: TU + MPI CBG)
- ***ANKA Tomography***
  Ultra Fast Tomography
- ***Nanoscience foundries and fine analysis (NFFA Europe)***
  EU
- ***Dariah & eCodicology & MASi***
  Arts & Humanities, ESFRI DARIAH EU + BMBF DARIAH DE, Metadata Management for Applied Sciences (MASi)

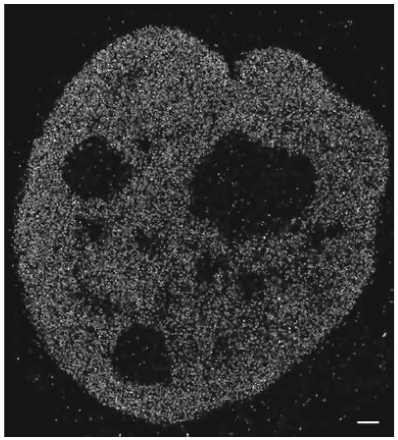
High Throughput Microscopy

- Discussions to automate standard tasks via UNICORE workflow engine
- Work on prototype HPC integration of KNIME desktop workflow system via UNICORE data oriented processing
- Evaluation of integrating microscopy use case with metadata project MASi
- Future:
  - Continuing work on HPC KNIME integration
  - Towards MASi prototype integration
  - Towards a generic provenance model?



Light Optical Nanoscopy



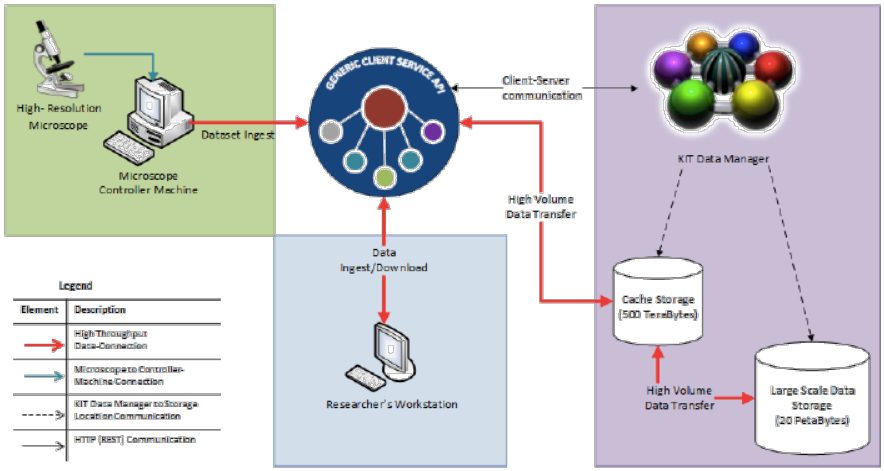


Open Reference Data Repository to manage large datasets (approx. 100 TB) for membrane and nucleus architecture on the nano-scale: storage, curation, reuse:

High performance data transfer client for data ingest and access installed on data acquisition system in Heidelberg

Metadata

- Metadata model for modeling ***provenance***
- Automated metadata extraction and model of the context metadata
- Web-portal enabling
  - discovery of datasets
  - sharing & referencing of datasets
  - executing scientific workflows

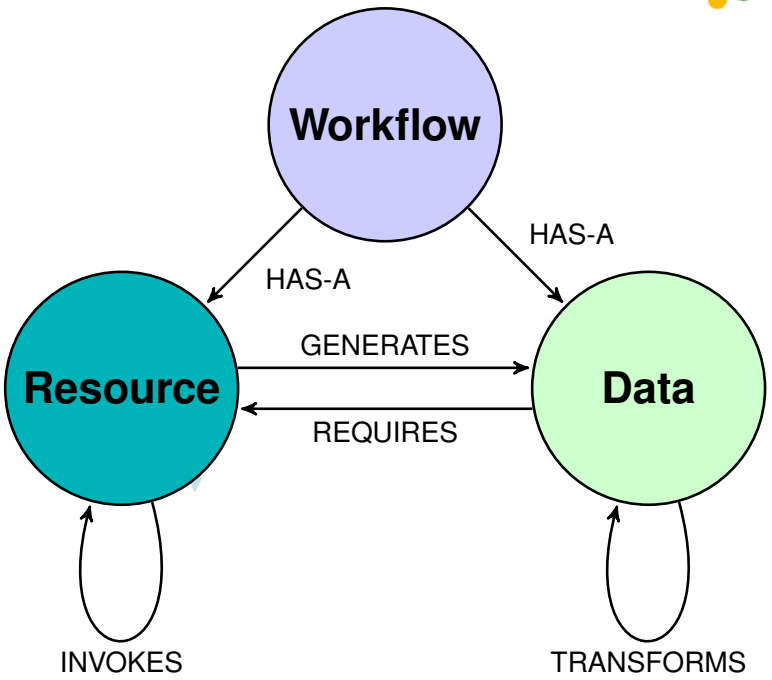


Impact: *Open Reference Data Repository*

LSDMA

Future developments

- RDF based metadata store for storing provenance metadata (OPM)
- Managing heterogeneous metadata, e.g. static vs. dynamic metadata
- Generic metadata API for managing heterogeneous metadata



RDA
RESEARCH DATA ALLIANCE

RDA Use Case:

- IG Data Fabric
- IG Metadata + WGs
- IG Research Data Provenance
- IG Repository Platforms

eCodicology

DARIAH-DE

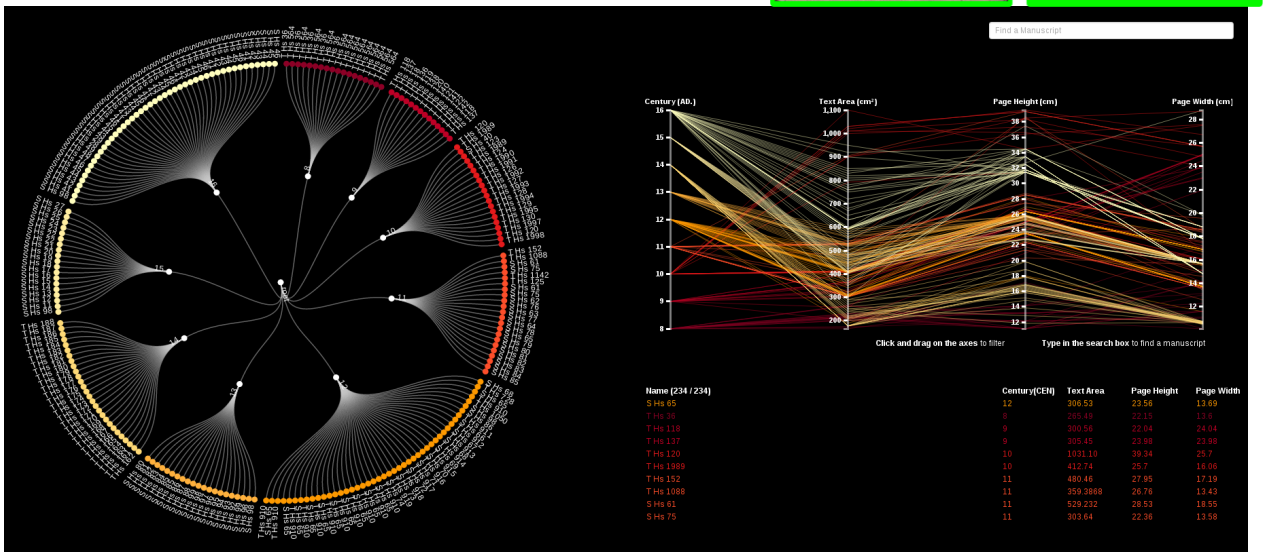
LSDMA

CodiLab:

- Integrating manual and automatic image annotation (+ vocabularies)
- Validation of automatic annotations

CodiVis

- Interactive data visualization of layout features
- Discovery of correlations



→ ***Visual analytics: new approach in humanities research***

→ ***Generally applicaple***

Historisch Kulturwissenschaftliche Informationsverarbeitung

Preservation

LSDMA
DARIAH-DE

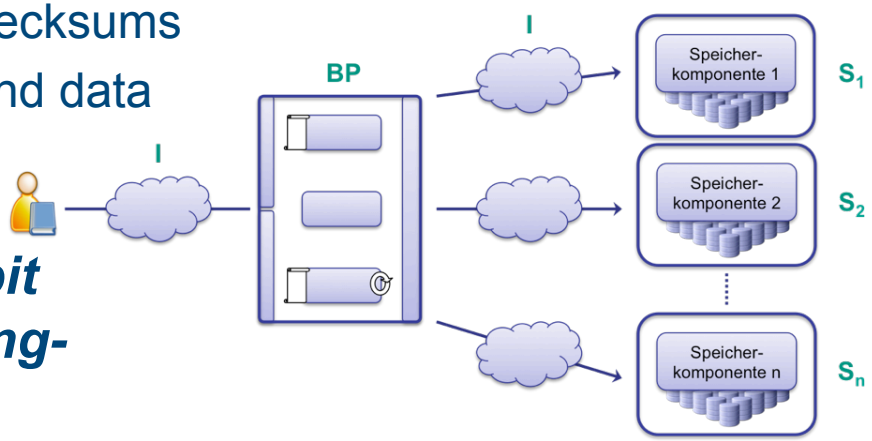
Investigation of reliability of bit preservation architectures

- Need for ***reliable metrics*** to quantify architectures
- Need for ***recommendations*** for fitting ***strategies***

Undetectable Error Rate:

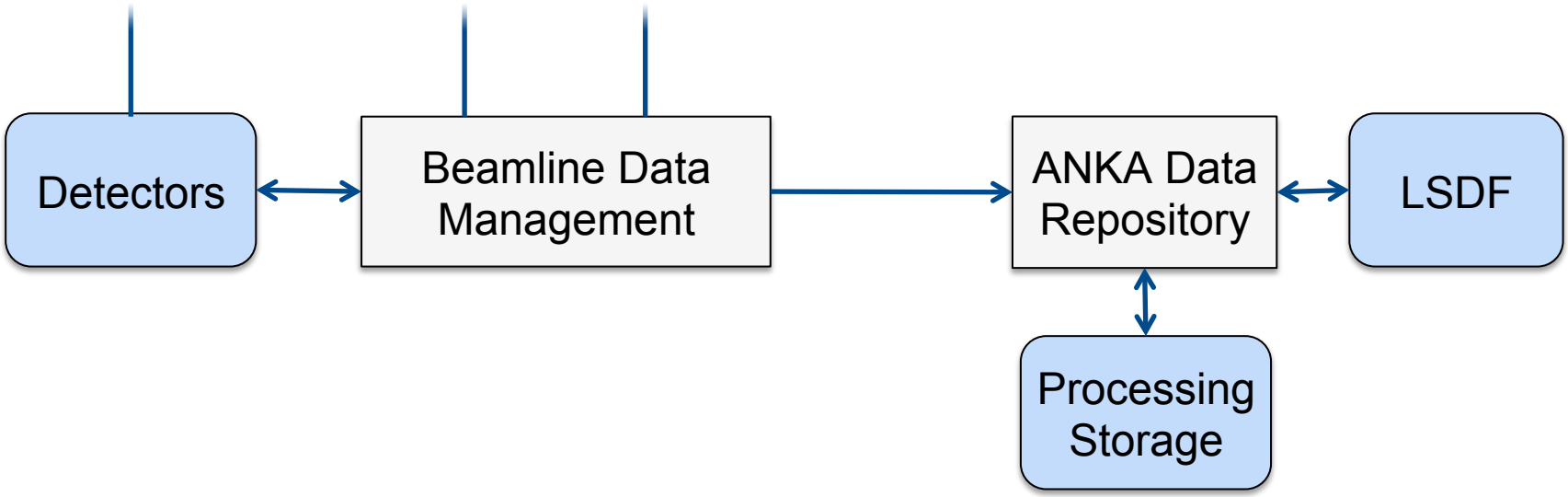
Probability P_{uError} – a bitstream contains undetected errors despite of actions to ensure data integrity

- Modeling + evaluation of replication scenarios
- Impact quantification of e.g. replication, checksums
- Method for classification of architectures and data allowing aligned bit preservation strategies



→ ***New approach for quantification of bit preservation architectures allowing long-term archiving***

ANKA Data Management & Repository



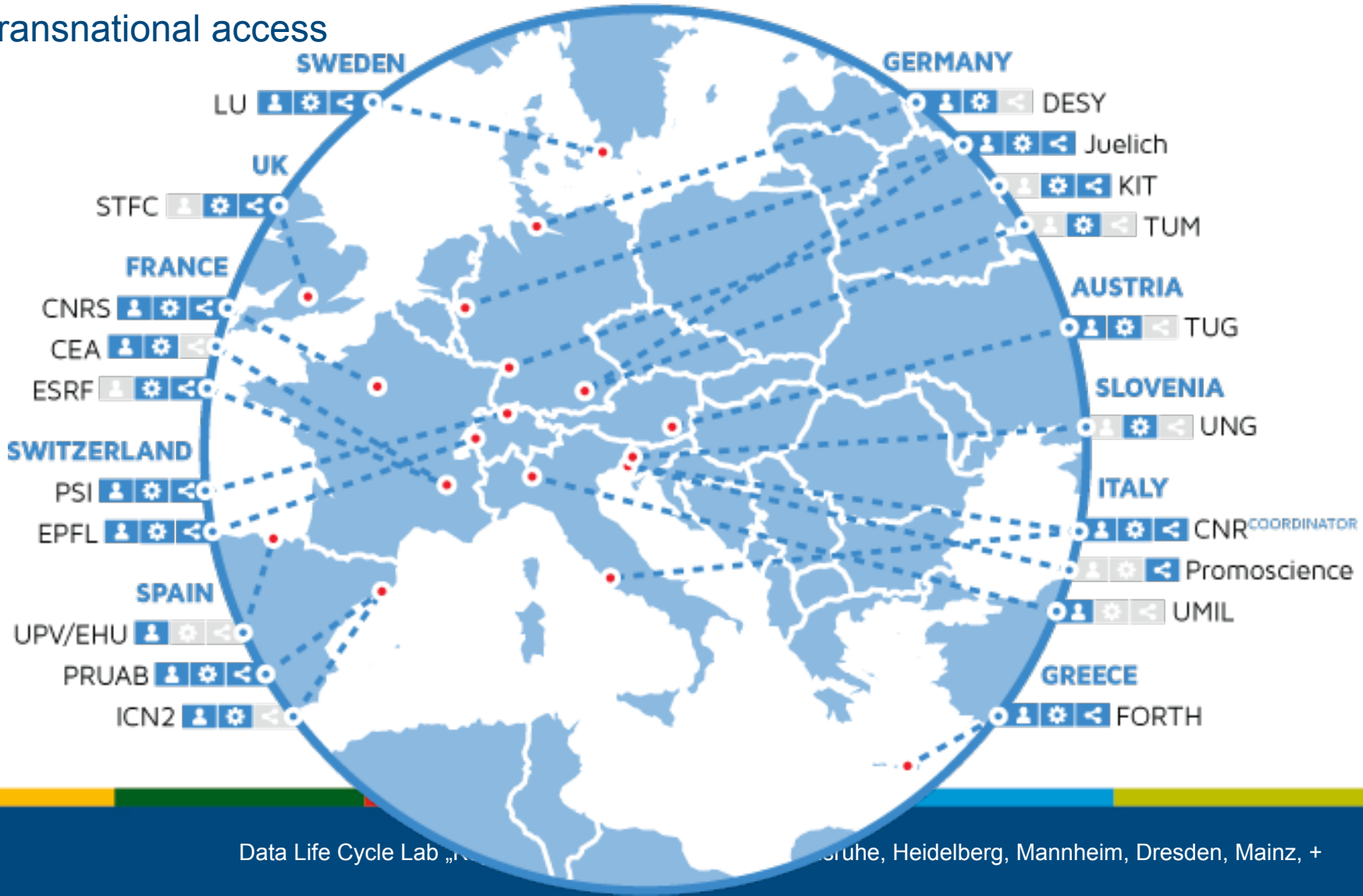
- Handover of the data management components

→ **In production data workflow for tomography data**

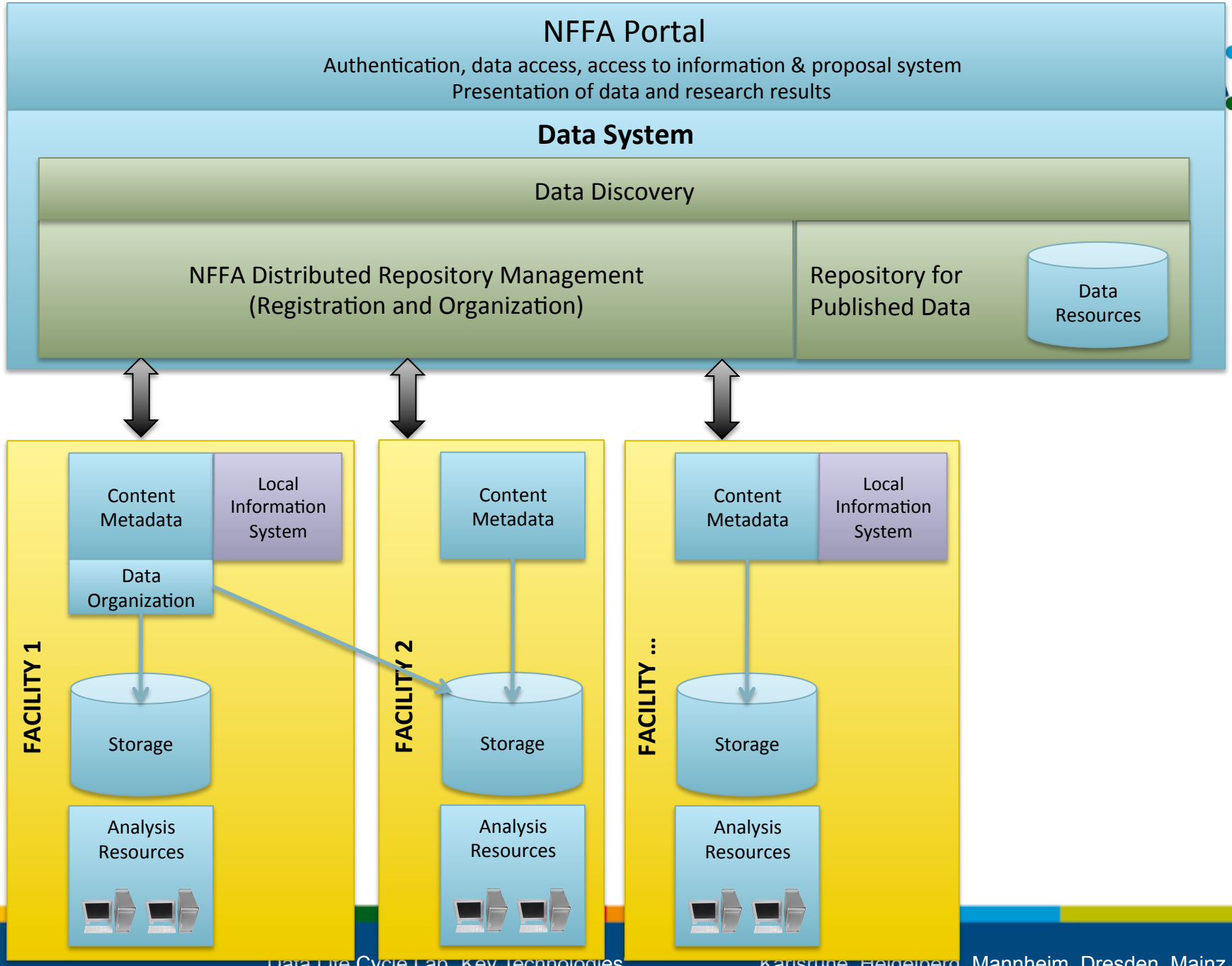
Nanoscience foundries and fine analysis (NFFA Europe)

LSDMA

European infrastructure
for transnational access



Data Life Cycle Lab „ruhe, Heidelberg, Mannheim, Dresden, Mainz, +


NFFA Portal
Authentication, data access, access to information & proposal system
Presentation of data and research results
Data System
Data Discovery
NFFA Distributed Repository Management
(Registration and Organization)
Repository for
Published Data
Data
Resources
Content
Metadata
Local
Information
System
Data
Organization
Storage
Analysis
Resources
FACILITY 1
Content
Metadata
Storage
Analysis
Resources
FACILITY 2
Content
Metadata
Local
Information
System
Storage
Analysis
Resources
FACILITY ...

Adoption of Existing Technologies

- **LSDMA**
  - Repository technologies, KIT Data Manager
- **DARIAH and EPIC**
  - PIDs,
  - Bit Preservation API
- **Research Data Alliance**
  - IG Metadata,
  - WG Data Type Registry,
  - WG Practical Policies
- **MASi**
  - Metadata infrastructure
- **PANdata**
  - iCAT
- **EUDAT**
  - B2Share

- IG Data Publication
- WG Metadata Standards Directory

Publications

- G. Nienhaus et al.: “An ensemble-averaged digital model of zebrafish embryo development based on light-sheet microscopy with single-cell resolution“, **Scientific Reports Nature**, 2015, DOI: 10.1038/srep08601
- X. Yang et al.: “TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray in vivo data“, **Optics Express**, 2015, DOI: 10.1364/OE.23.005368
- A. Prabhune, et al.: “An Optimized Generic Client Service API for Managing Large Datasets within a Data Repository”, **Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications**, 2015, BigDataService 2015, San Francisco: Paper ID 33
- Y. Zhang, et al.: “Radiation induced chromatin conformation changes analysed by fluorescent localization microscopy, statistical physics, and graph theory“, **PLoS ONE**, 2015, PloS One10 (6): e13636, doi: 10.1371/journal.pone.0013636
- S. Herres-Pawlis,et al.: “Quantum chemical meta-workflows in MoSGrid“, **Concurrency and Computation: Practice and Experience**, 2015, 27, 344-357
- R. Grunzke and R. Müller-Pfefferkorn: “Certificate-free User-friendly HPC Access with UNICORE“, **UNICORE Summit 2014**, Proceedings
- R. Grunzke, S. Gesing, R. Jäkel and W. E. Nagel: “Towards Generic Metadata Management in Distributed Science Gateway Infrastructures“, **IEEE/ACM CCGrid 2014** (14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing), 2014

Publications

- A. Hoffmann, S. Herres-Pawlis, L. de la Garza, R. Grunzke and J. Krüger: “Expansion of Quantum Chemical Metadata for Workflows in the MoSGrid Science Gateway“, **IWSG 2014** (6th International Workshop on Science Gateways), 2014
- S. Herres-Pawlis, et al.: “Meta-metaworkflows for Combining Quantum Chemistry and Molecular Dynamics in the MoSGrid Science Gateway“, **IWSG 2014** (6th International Workshop on Science Gateways), 2014
- S. Gesing, et al.: “A Science Gateway Tailored to the Molecular Simulation Community“, Book Chapter in “Science Gateways for Distributed Computing Infrastructures”, **Springer**, 2014
- S. Pyatykh and Hesser, J.: “Image Sensor Noise Parameter Estimation by Variance Stabilization and Normality Assessment“, IEEE Transactions on Image Processing, 2014, vol. 23, no. 9, p. 3990,3998,
- S. Pyatykh and Hesser, J., “Salt and pepper noise removal in binary images using image block prior probabilities“, Journal of Visual Communication and Image Representation, 2014, vol. 25, pp. 748-754
- G. A. McGilvary, M. Atkinson, S. Gesing, A. Aguilera, R. Grunzke and E. Sciacca: Enhanced Usability of Managing Workflows in an Industrial Data Gateway, Interoperable Infrastructures for Interdisciplinary Big Data Sciences (IT4RIs 15), accepted.
- R. Grunzke, J. Krüger, S. Gesing, S. Herres-Pawlis, A. Hoffemann, A. Aguilera and W. E. Nagel: Managing Complexity in Distributed Data Life Cycles Enhancing Scientific Discovery, 11th IEEE International Conference on eScience, accepted.

Publications

- A. Aguilera, R. Grunzke, U. Markwardt, D. Habich, D. Schollbach and J. Garcke: Towards an Industry Data Gateway: An Integrated Platform for the Analysis of Wind Turbine Data, Science Gateways (IWSG), 7th International Workshop on, 2015, 62-66.
- S. Gesing, J. Krüger, R. Dooley, R. Grunzke, M. Pierce, S. Herres-Pawlis and A. Hoffmann: Science Gateways – Leveraging Modeling and Simulations in HPC Infrastructures via Increased Usability, The International Conference on High Performance Computing & Simulation (HPCS 2015), accepted.
- S. Gesing, J. Kruger, R. Grunzke, S. Herres-Pawlis and A. Hoffmann: Challenges and Modifications for Creating a MoSGrid Science Gateway for US and European Infrastructures, Science Gateways (IWSG), 7th International Workshop on, 2015, 73-79.
- S. Herres-Pawlis, A. Hoffmann, T. Rosener, J. Kruger, R. Grunzke and S. Gesing: Multi-layer Meta-metaworkflows for the Evaluation of Solvent and Dispersion Effects in Transition Metal Systems Using the MoSGrid Science Gateways, Science Gateways (IWSG), 2015 7th International Workshop on, 2015, 47-52.

Adoption of Existing Technologies

- **LSDMA**
  - Repository technologies, KIT Data Manager
- **DARIAH and EPIC**
  - PIDs,
  - Bit Preservation API
- **Research Data Alliance**
  - IG Metadata,
  - WG Data Type Registry,
  - WG Practical Policies
  - IG Data Publication
  - WG Metadata Standards Directory
- **MASi**
  - Metadata infrastructure
- **PANdata**
  - iCAT
- **EUDAT**
  - B2Share