

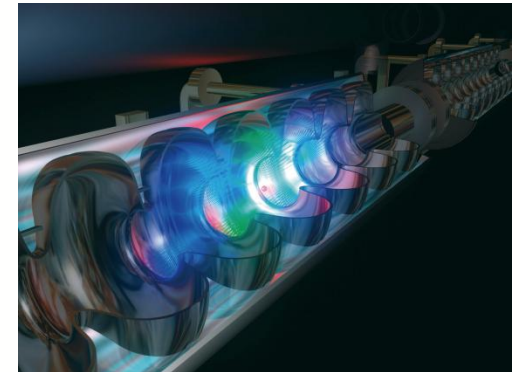
# Data Life Cycle Labs, a new concept to support data-intensive science

Christopher Jung, Achim Streit, Rainer Stotzka, Jörg Meyer, Silke Halstenberg, Fabian Rigoll, Ariel Garcia, Andreas Heiss, Jos van Wezel (all KIT), Kilian Schwarz (GSI), Martin Gasthuber (DESY), André Giesler (FZJ)

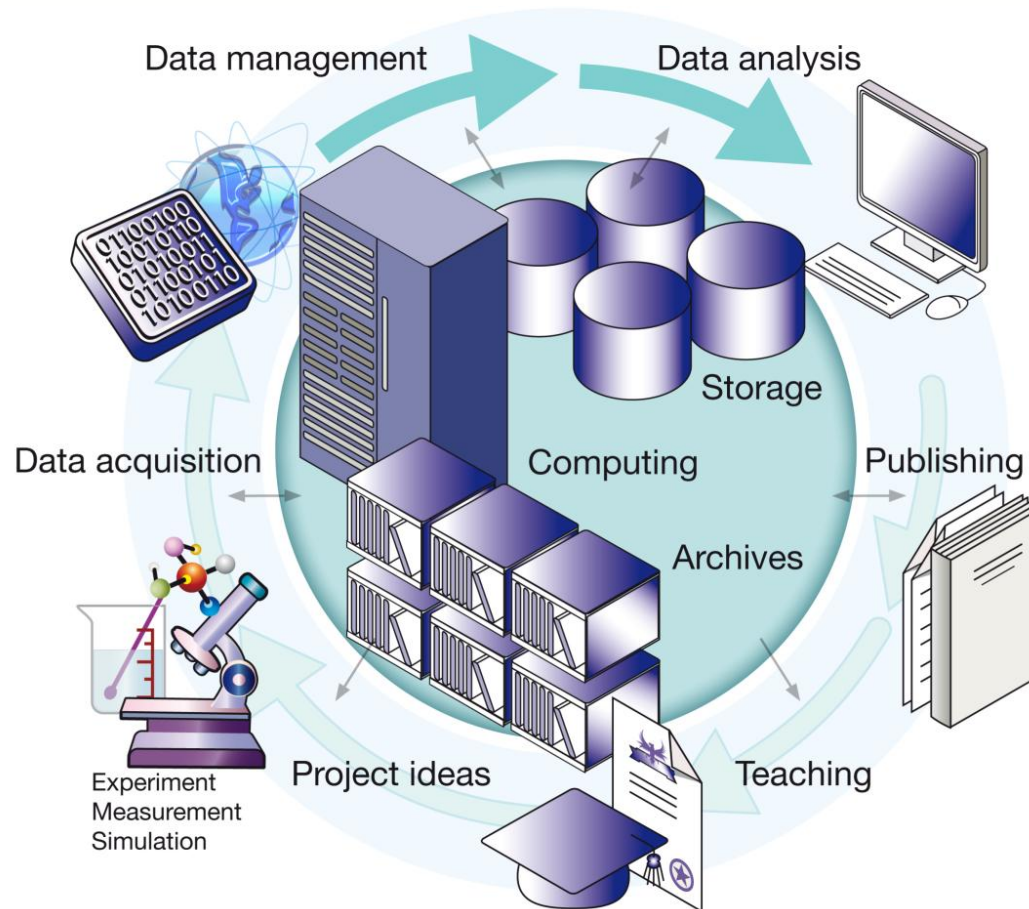


# Scientific Data are ...

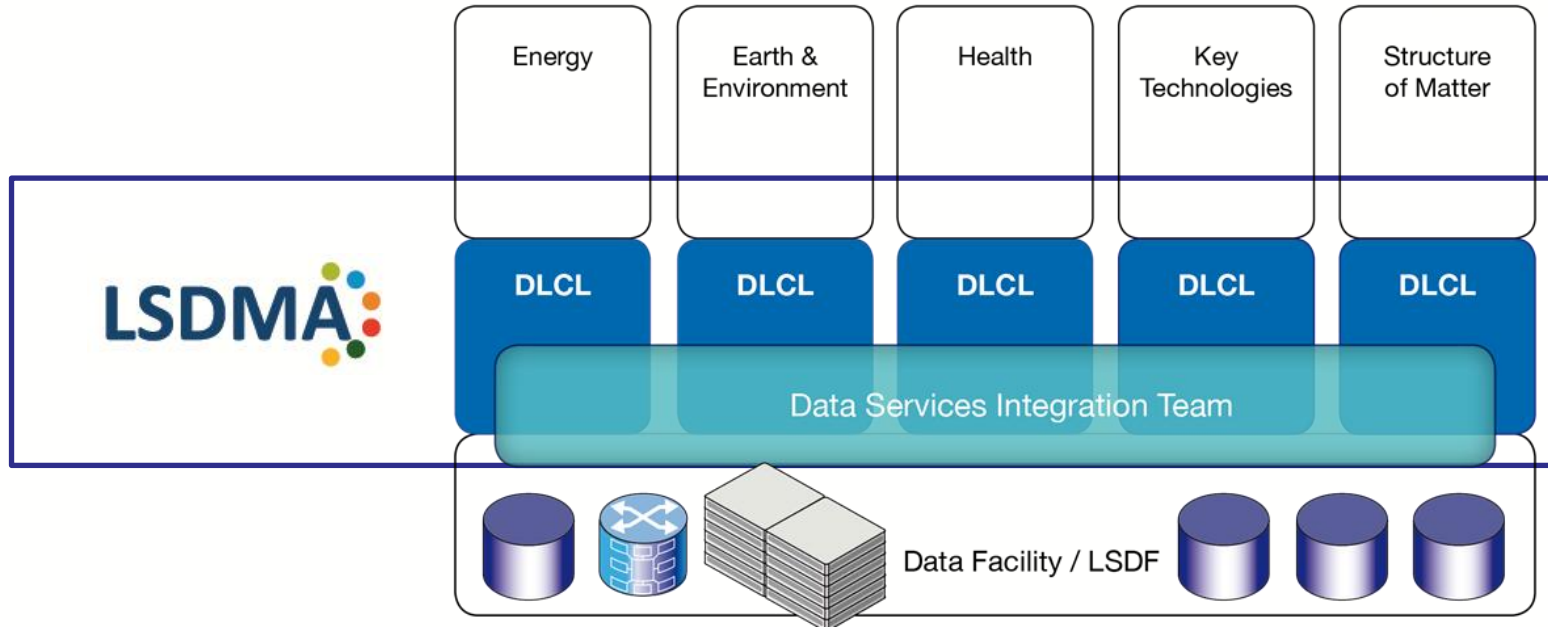
- ... collected and recorded at a rapidly increasing rate → **'data deluge'**
- ... analyzed by **distributed collaborations** and stored on **distributed resources**
- ... cannot be reproduced easily or at all → **valuable good**
- ... have to be **preserved**
- ... need to be described well → **metadata**
- ... usually collected by researchers from one field, but might be used by **multiple disciplines**
- ... analyzed by methods which scale poorly → **improve or replace analysis methods**



# Scientific Data Life Cycle



# LSDMA Structure



## Data Life Cycle Labs

Joint R&D with the scientific user communities:

- Optimization of the data life cycle
- Community-specific data analysis tools and services

## Data Services Integration Team

Generic methods R&D:

- Data analysis tools and services common to several DLCLs
- Interface between federated data infrastructures and DLCLs/communities

# Facts and Figures

- Helmholtz Portfolio Extension
- Initial project duration: 2012-2016
- Eleven partners:



RUPRECHT-KARLS-  
UNIVERSITÄT  
HEIDELBERG



ulm university universität

uulm



Universität Hamburg



- Project lead: KIT
- Sustainability: inclusion of activities into respective Helmholtz program-oriented funding in 2015

# Highlights of current DLCL activities (I)

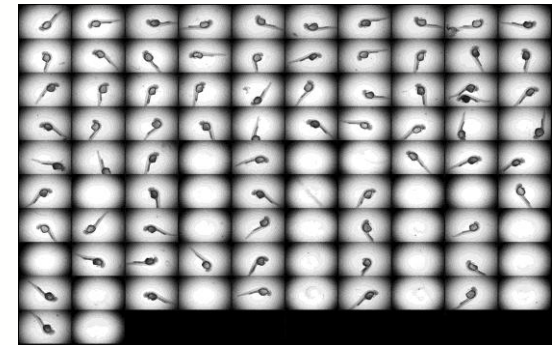
## DLCL Energy:

- Life cycle analysis in various energy data projects
- Analyzing stereoscopic satellite for estimating the efficiency of solar energy with Hadoop
- Transferring local analysis to clusters



## DLCL Key Technologies:

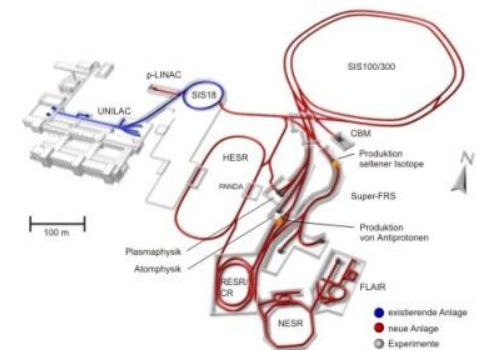
- Identification and development of joint tools and services for several communities
- Optimization of tomographical reconstruction using data-intensive computing
- On-server visualization for high throughput microscopy



# Highlights of current DLCL activities (II)

## DLCL Structure of Matter:

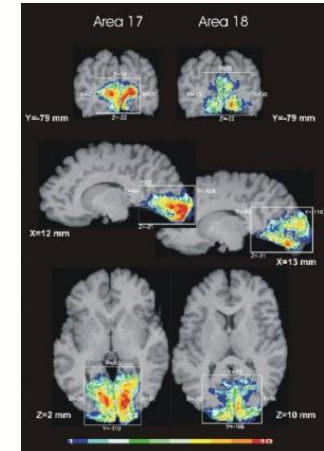
- Development of a portal for PETRA-III data
- IDM for PETRA-III experimentators
- Optimizing dCache for efficiently storing small files on tape
- Determining specifications of an archive system for photon science and high energy physics
- Determining the computing requirements for FAIR data analysis
- Modular concept for representing of detector geometry as a basis for simulation



# Highlights of current DLCL activities (III)

## DLCL Health:

- Workflow support for data-intensive parameter studies
- Data generation, transfer and storage for the Brain Big Data Collaboration with University of Düsseldorf
- Efficient metadata administration and indexing



## DLCL Earth&Environment:

- MongoDB for data and metadata of meteorologic satellite data
- Local data catalog for meteorologic simulation data downloaded from ECMWF
- Replacing a MySQL databank for satellite metadata
- Data replication in an iRODS federation





## Work packages

- Federated ID management
- Federated data access
- Metadata catalogs and repositories
- Archive Services
- Monitoring, modeling, optimization
- Data intensive computing

Topics within work packages driven by joint needs of research communities.

# LSDMA Events



- Annual international Big Data symposium at KIT



- Annual community forum



- Topical workshops, e.g. on identity management



# Summary and outlook



- Sustainable r&d project, focusing on joint research with communities
- Data Life Cycle Labs for five research fields of the Helmholtz Association
- Partners: four Helmholtz centers, six universities, German climate research center
- Broad spectrum of current activities in DLCLs
- Generic r&d in Data Services Integration Team, driven by joint needs of the research communities.
- 2<sup>nd</sup> International Symposium on September 24<sup>th</sup> at KIT