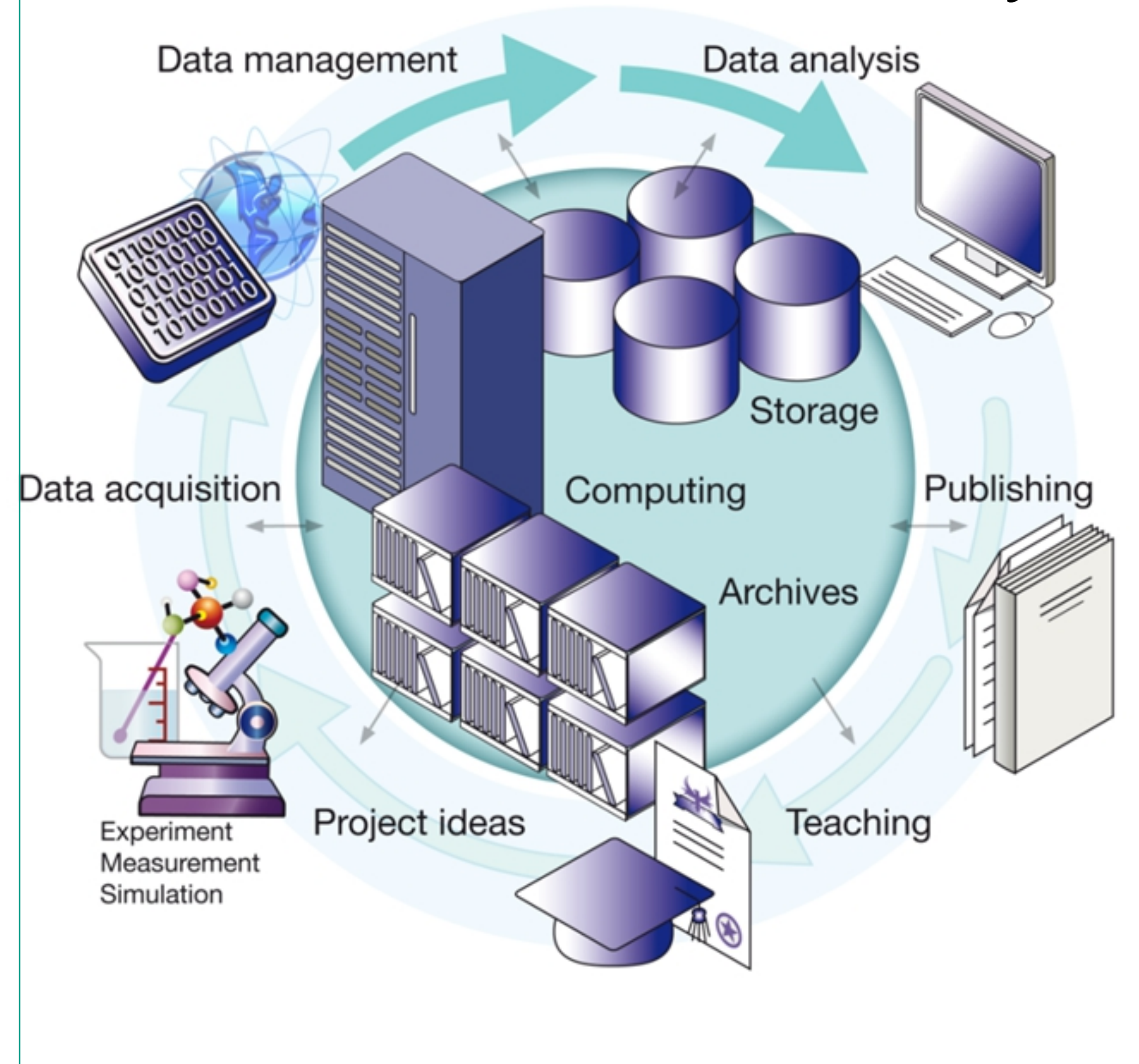


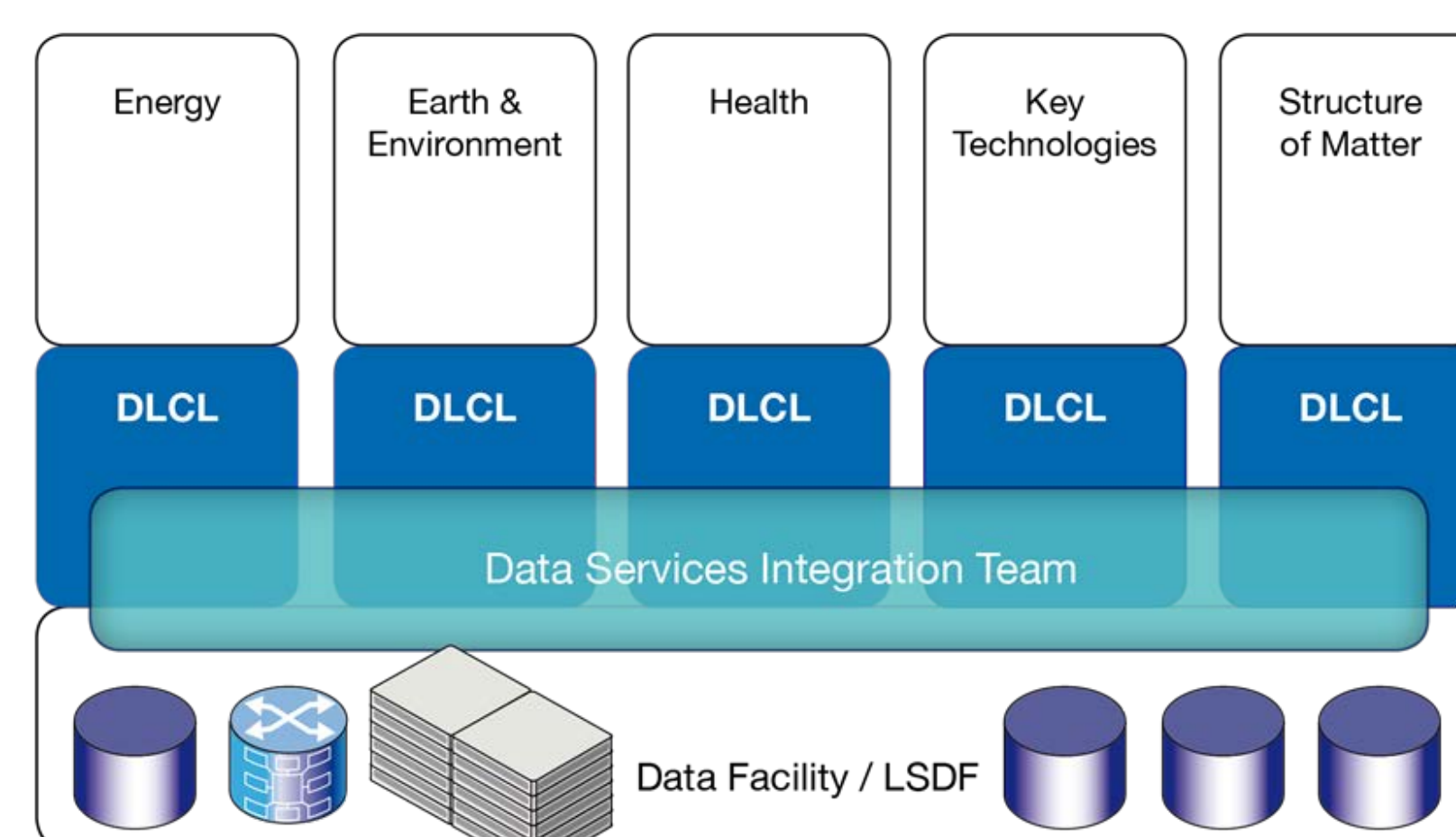
Progress in Multi-Disciplinary Data Life Cycle Management

Marcus Hardt, Christopher Jung, Jörg Meyer, Ajinkya Prabhune, Fabian Rigoll, Rainer Stotzka, Achim Streit, KIT; André Giesler, FZ Jülich; Martin Gasthuber, DESY Hamburg; Kilian Schwarz, GSI Darmstadt

Data in the Scientific Life Cycle

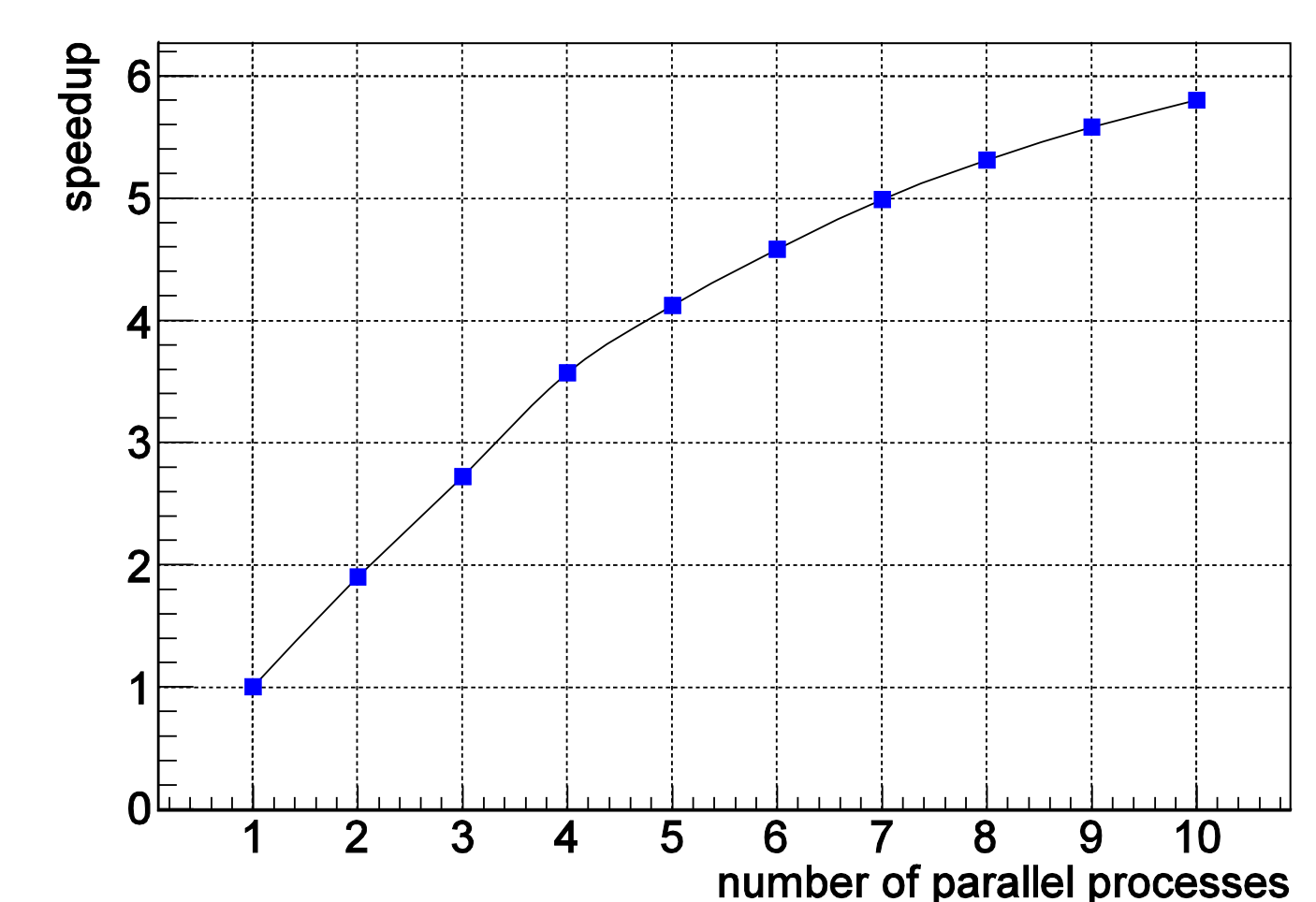


Large-Scale Data Management and Analysis



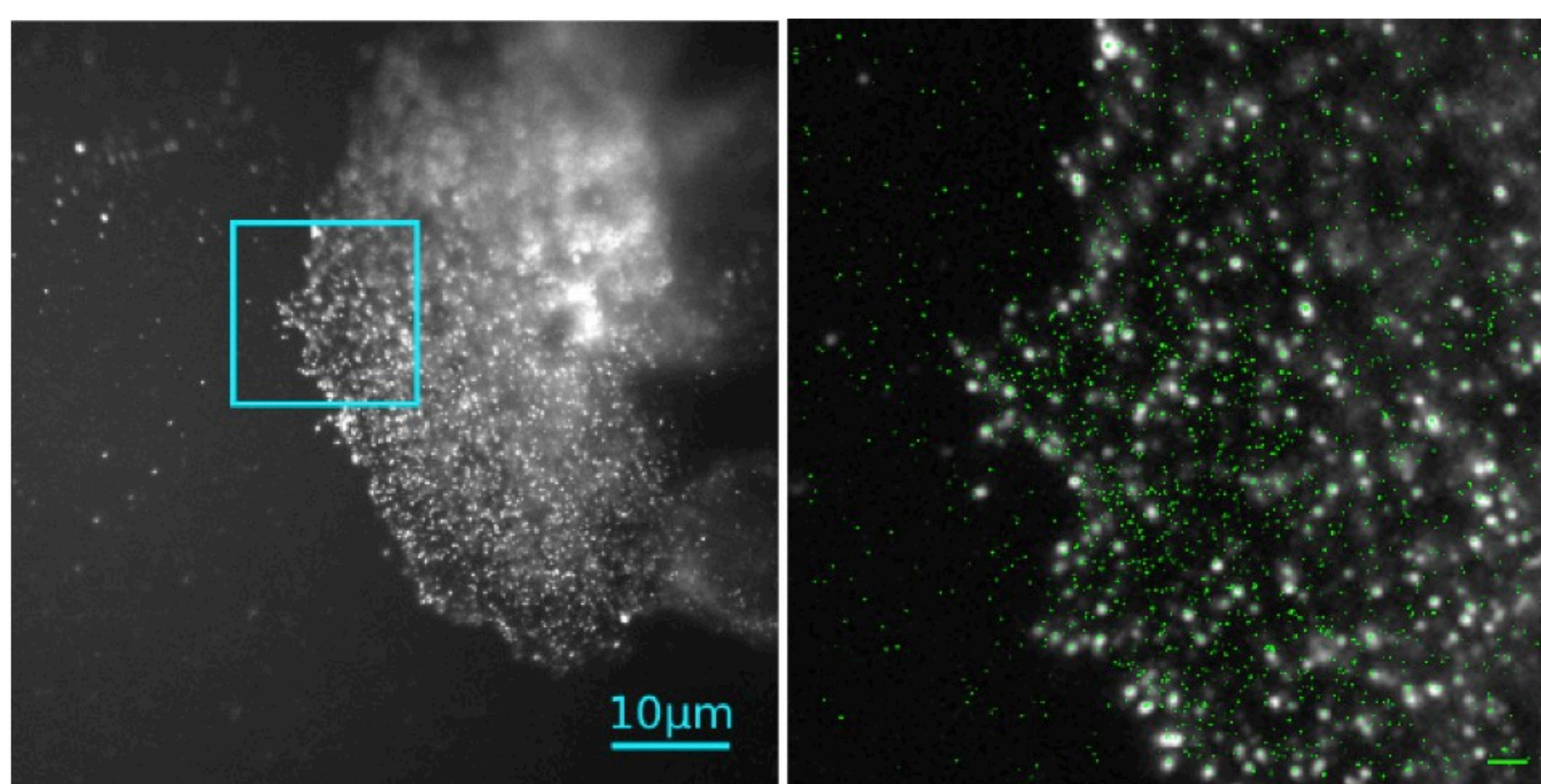
- Portfolio Extension and Cross-Sectional Activity of the German Helmholtz Association
- Activities incorporated into the third period of the program-oriented funding

DLCL Earth and Environment



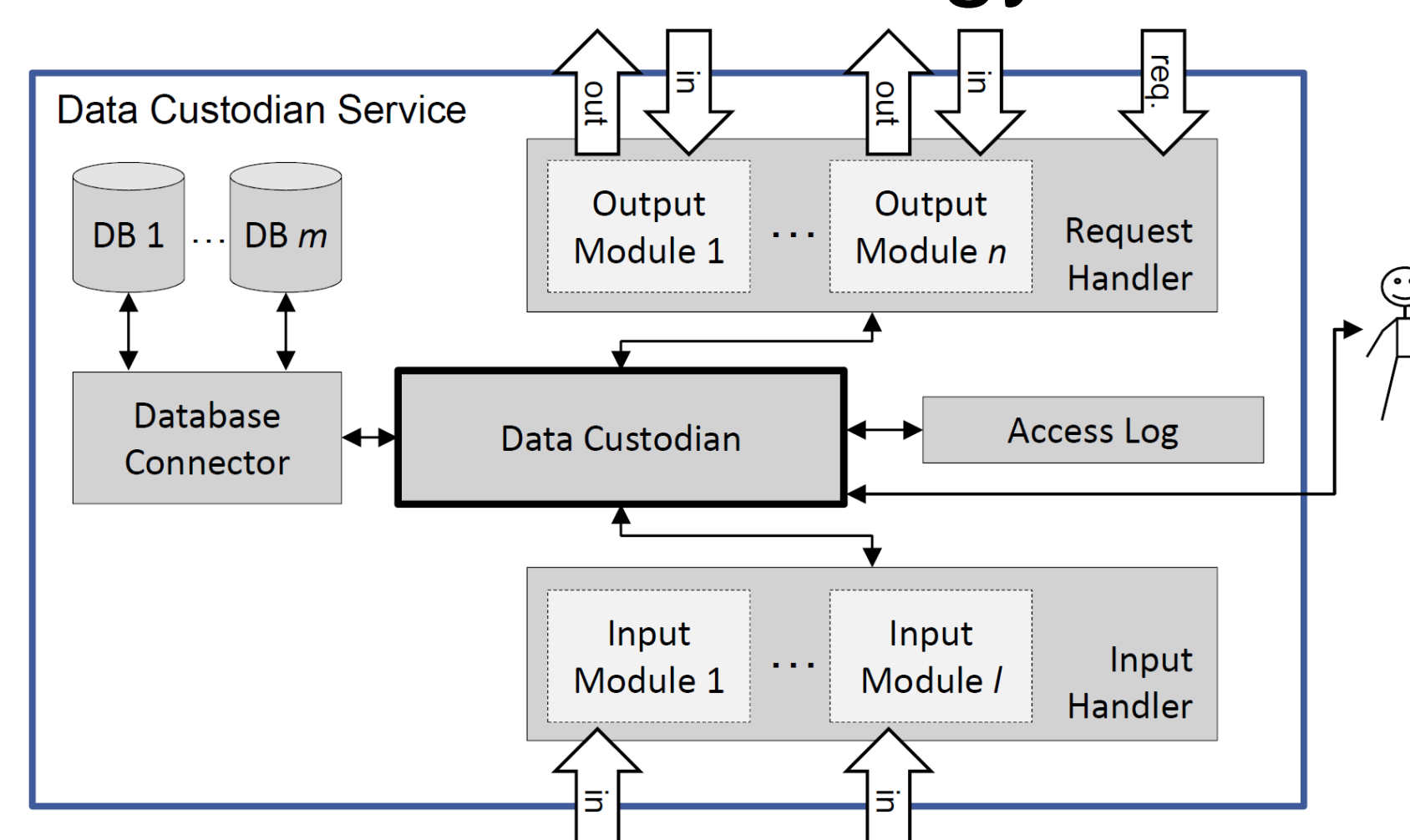
- Speeding up geo-matching for data from different instruments using MongoDB

DLCL Key Technologies



- KIT Data Manager is a repository system for scientific datasets, e.g from high-throughput microscopy
- Handles sharing, referencing, curation, data transfer and data discovery based on metadata

DLCL Energy

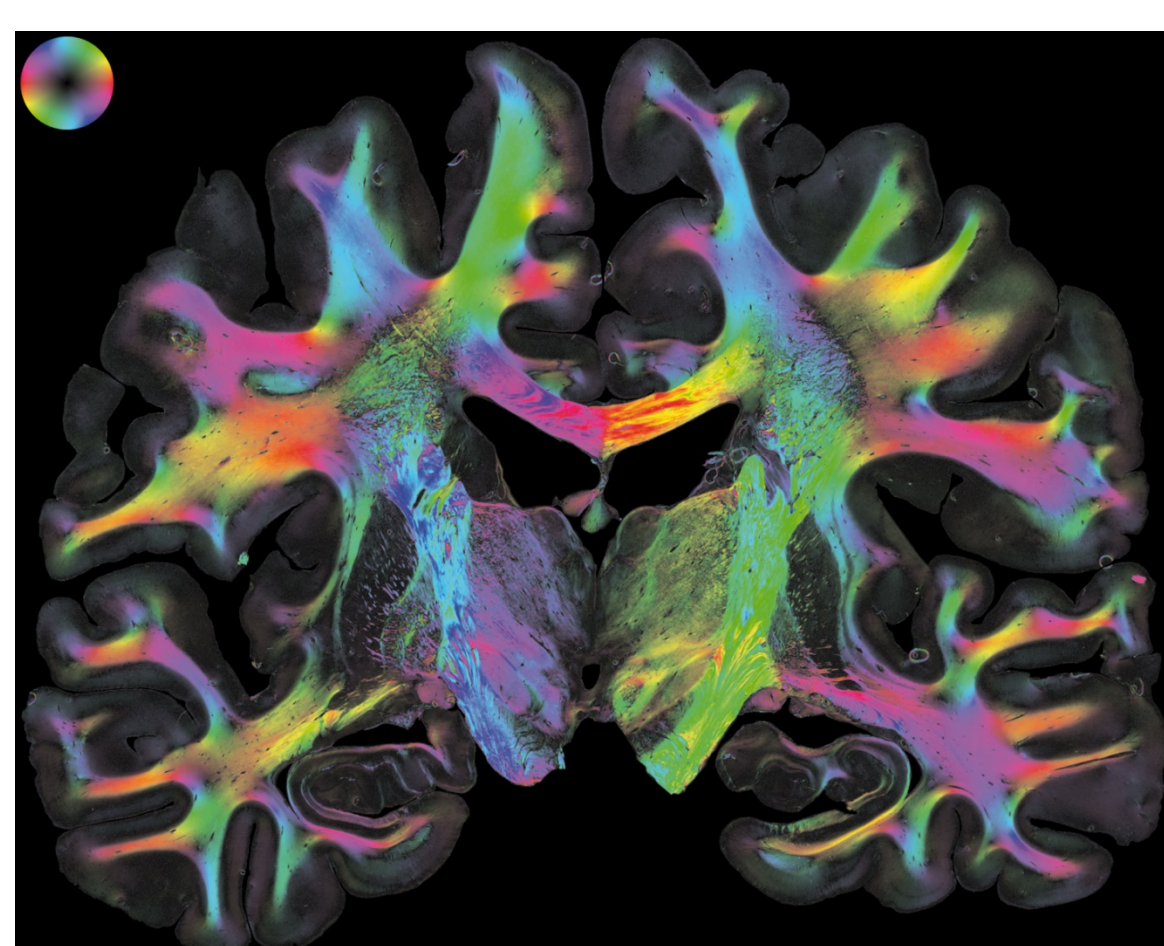


- Energy consumption is most often personal data needing special protection
- Data custodian service allows user-defined rules for data sharing data for each interested 3rd party (potentially w/quality restrictions)

Data Services Integration Team

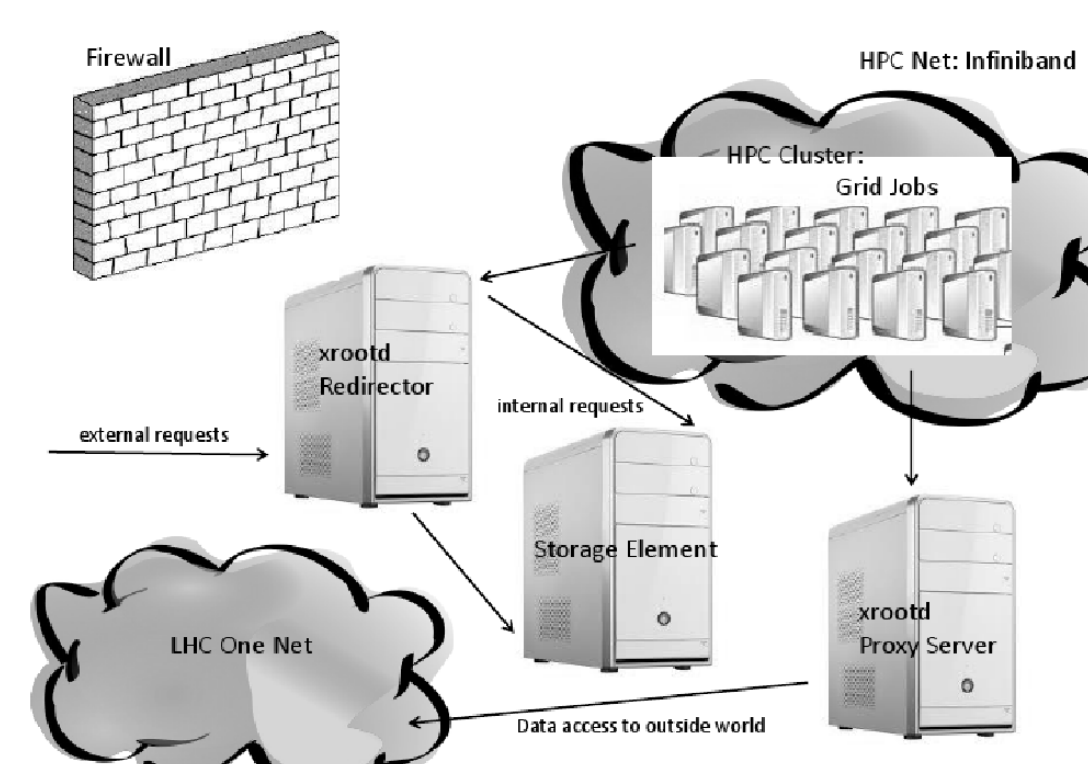
- Provide migration paths from X.509 certificates to other authentication infrastructures such as SAML or OpenIDConnect via two approaches
 - User visits the SLCS portal and authenticates via SAML at his home institutions identity-provider (IdP); DFN SLCS CA is accredited within the IGTF trust framework and can issue certificates for use in LCG
 - Extend LDAP as one of the core components that handle local authentication of users

DLCL Health



- Three Dimensional Polarized Light Imaging allows the study of brain regions with resolution at sub-millimeter scale
- Calibration, component analysis, enhanced analysis, stitching and segmentation integrated in a UNICORE workflow, resulting in speedup of $O(100)$ and easy reproducibility

DLCL Structure of Matter



- Inclusion of local HPC resources into a distributed computing environment for FAIR; main elements: xrootd redirector as well as xrootd forward proxy server
- *For Photon Science activities : talk 498 (Thursday, 9:15 in C209)*

Conclusions and Lessons Learned

- Big Data challenges often specific to scientific discipline
- Scientists want and need to focus on analysis, not data management
- Communities need more than just storage space
- Syncing domain-specific and generic R&D takes time
- Communities need evolution, not revolution

The authors wish to thank all people and institutions involved in LSDMA as well as the German Helmholtz Association for funding.