Developing the WLCG Tier-1 center GridKa as topical center in a multidisciplinary research environment
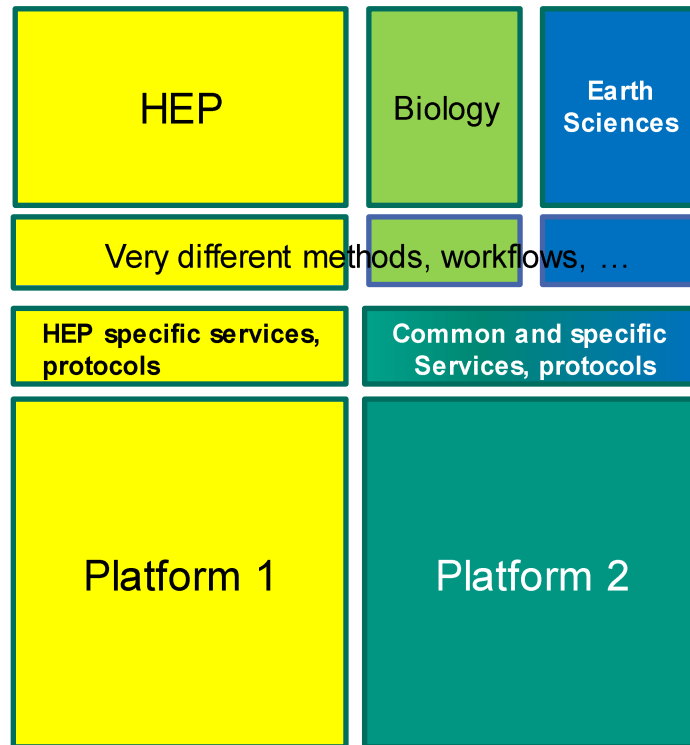
andreas.heiss@kit.edu, andreas.petzold@kit.edu
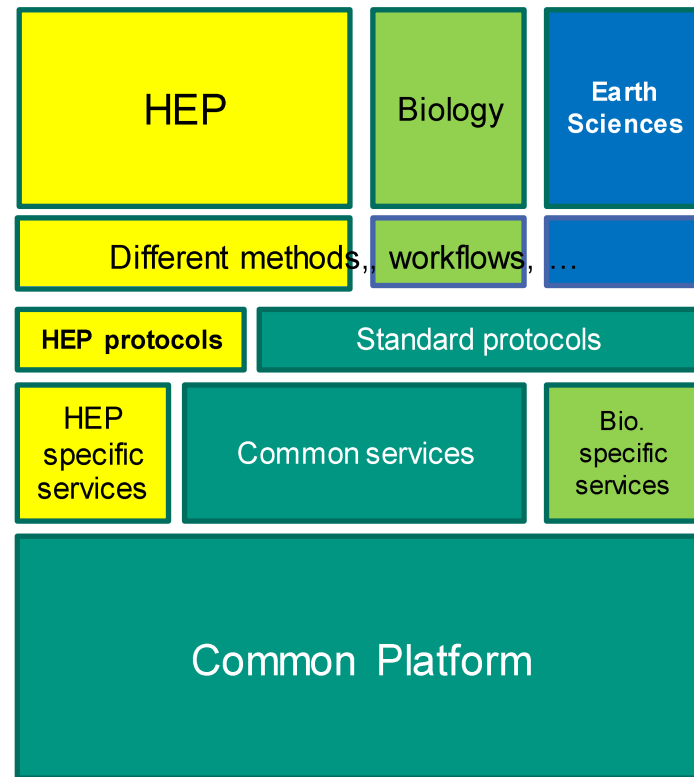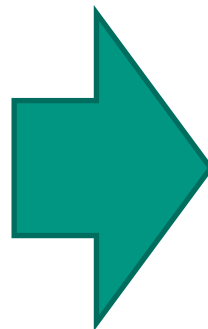
Steinbuch Centre for Computing

Motivation

- Flat budgets, increasing costs for energy, personnel, …

- Increasing competition by commercial cloud providers

- HEP computing requirements (# PBs etc.) not unique any more. Other fields of research catching up.

- Funding agencies hesitant to fund HEP specific computing infrastructures and push for more cooperation between different research fields.
    - => National and EC funded projects and infrastructures

| HEP | Biology | Earth Sciences |

Very different methods, workflows, …

HEP specific services, protocols

Common and specific Services, protocols

| Platform 1 | Platform 2 |

Steinbuch Centre for Computing

Motivation

- Flat budgets, increasing costs for energy, personnel, …

- Increasing competition by commercial cloud providers

- HEP computing requirements (# PBs etc.) not unique any more. Other fields of research catching up.

- Funding agencies hesitant to fund HEP specific computing infrastructures and push for more cooperation between different research fields.
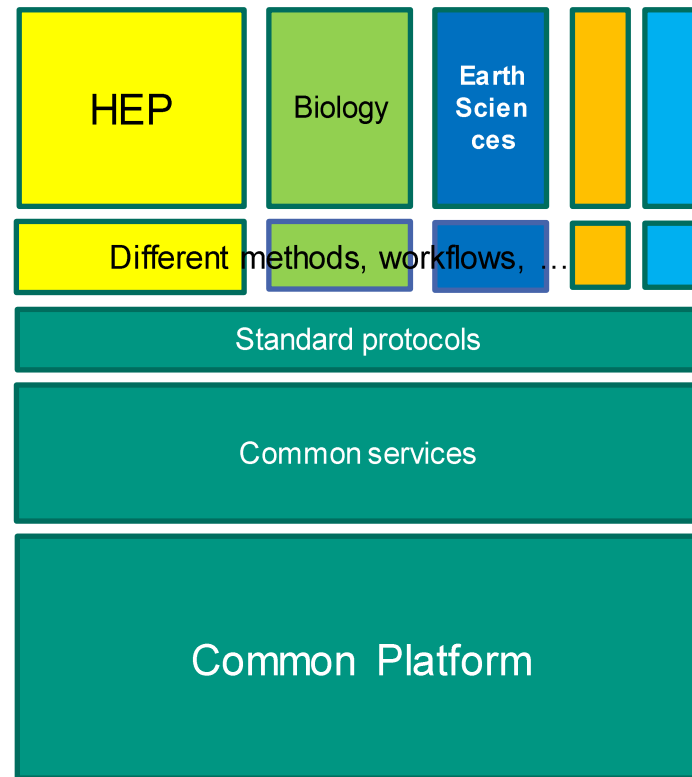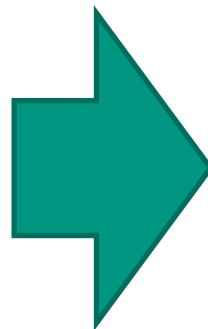
  => National and EC funded projects and infrastructures

CHEP 2016, A. Heiss, A. Petzold

Motivation

- Flat budgets, increasing costs for energy, personnel, …
- Increasing competition by commercial cloud providers
- HEP computing requirements (# PBs etc.) not unique any more. Other fields of research catching up.
- Funding agencies hesitant to fund HEP specific computing infrastructures and push for more cooperation between different research fields.
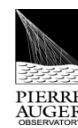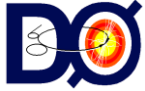 - => National and EC funded projects and infrastructures

HEP

Biology

Earth Sciences

Different methods, workflows, …

Standard protocols

Common services

Common Platform

KIT data center for High Energy Physics

- Started in 2003 as a (Grid) computing center for High Energy Physics

- Today among the largest WLCG Tier-1 centers, providing ~15% of the total WLCG Tier-1 resources.
  - 2016: 11k CPU cores (168 kHS'06), 16 PB disk, 20 PB tape
  - The cornerstone of "Big Data" research and infrastructures at KIT (and Helmholtz)
    - => Helmholtz research program "Supercomputing and BigData" at KIT
    - => Large Scale Data Facility at KIT

Large Scale Data Facility @ KIT

- Started 2008 with a 150 TB storage system to provide central storage for KIT researchers
  - High throughput microscopy
  - Climate research
  - Synchrotron radiation techniques (imaging, spectroscopy, …)
- 2009: First KIT-external research group uses LSDF

- 2016:
  - > 5 PB disk and > 6 PB tape storage and growing
  - Additional storage services for various users (groups) within the state of Baden-Württemberg and beyond.

Steinbuch Centre for Computing
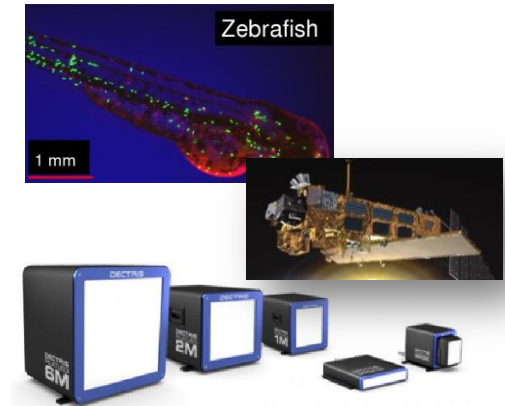
Integrating GridKa and LSDF

- Until 2013: Different teams, methods, tools, systems
- Since 2013:
  - Operations manager responsible for GridKa and LSDF
  - Common personnel (for certain systems and services, where reasonable)
  - Similar working procedures
  - Common operations tools
  - Common infrastructures

- Starting 2016 / 2017
  - One funding source
  - Common procurements
  - Similar systems (later: common systems)

→
  - Operations and expert knowledge shared between larger number of people
  - Consolidation allows for more staff for community support and R&D
  - Profit from economy of scale

Steinbuch Centre for Computing

Integrating GridKa and LSDF: 2016

16PB disk

GridKa compute farm

1&1 cloud

bwHPC

20+6 PB tape

TSM

See CHEP talk #400:
T. Hauth, "On-demand provisioning of HEP compute resources on cloud sites and shared HPC centers"
Wed. 12:30 pm Track 3

6PB disk (IBM Sonas)

- 56 node
- Interactive nodes
- NFS to HPC systems

Steinbuch Centre for Computing

Integrating GridKa and LSDF

2017 **2018** long-term

**Similar storage hardware
GPFS as 'virtualisation layer'
(common procurement)**

Common HPSS
tape backend

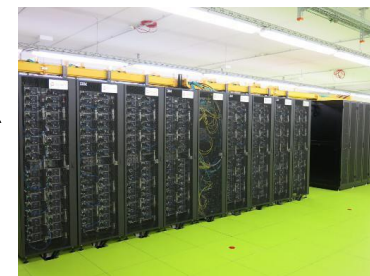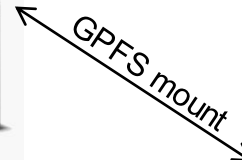Further common services:
AAI, data management, …
HDF

LSDF

20+ PB disk
- GPFS
- 2017 pledge increase
- Replacement
  of old systems

9 PB disk

GPFS mount

GridKa compute
Farm
-PoC: Extension to cloud

HELIX
NEBULA
THESCIENCECLOUD

amazon
web services
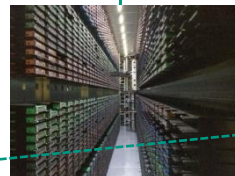
TSM → **HPSS**
High Performance Storage System

ForHLR Petaflop-System

SCC Steinbuch Centre for Computing

Integrating GridKa and LSDF

CHEP 2016, A. Heiss, A. Petzold

Steinbuch Centre for Computing
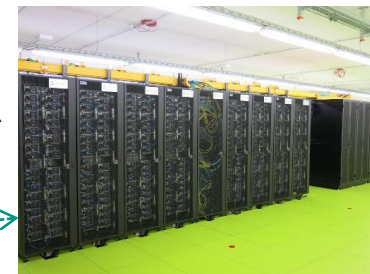
- Federated research data infrastructure for Germany.

- Funding of data and computing infrastructure at 6 centers in 2017 - 2021
  - **Includes (significant part of) funding of the WLCG Tier-1 resources at KIT**

- Polar Ice Sheets
- Virtual Observatory for Polar and Marine Research Data

DESY Tier-2
- Particle Physics Experiments
- Photon Science at Petra III, Flash & Flash 2
- Helmholtz Beamline @ EU-XFEL

- Genome Research
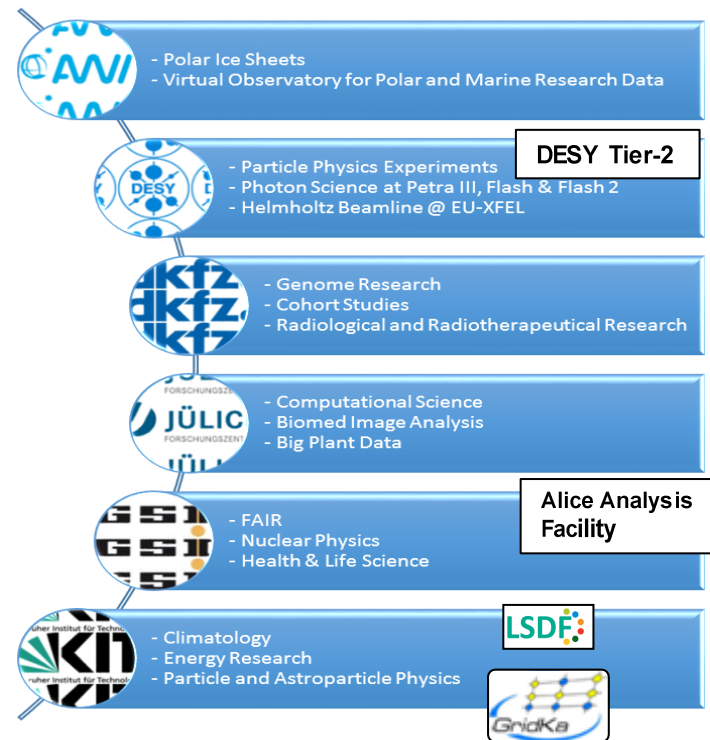- Cohort Studies
- Radiological and Radiotherapeutical Research

- Computational Science
- Biomed Image Analysis
- Big Plant Data

Alice Analysis Facility
- FAIR
- Nuclear Physics
- Health & Life Science

LSDF
- Climatology
- Energy Research
- Particle and Astroparticle Physics

GridKa

SCC Steinbuch Centre for Computing

- Planned services include:
  - Common AAI
  - Data transfer
  - Backup and archival
  - Software distribution
  - Metadata services
  - Compute cloud
  - Storage federation
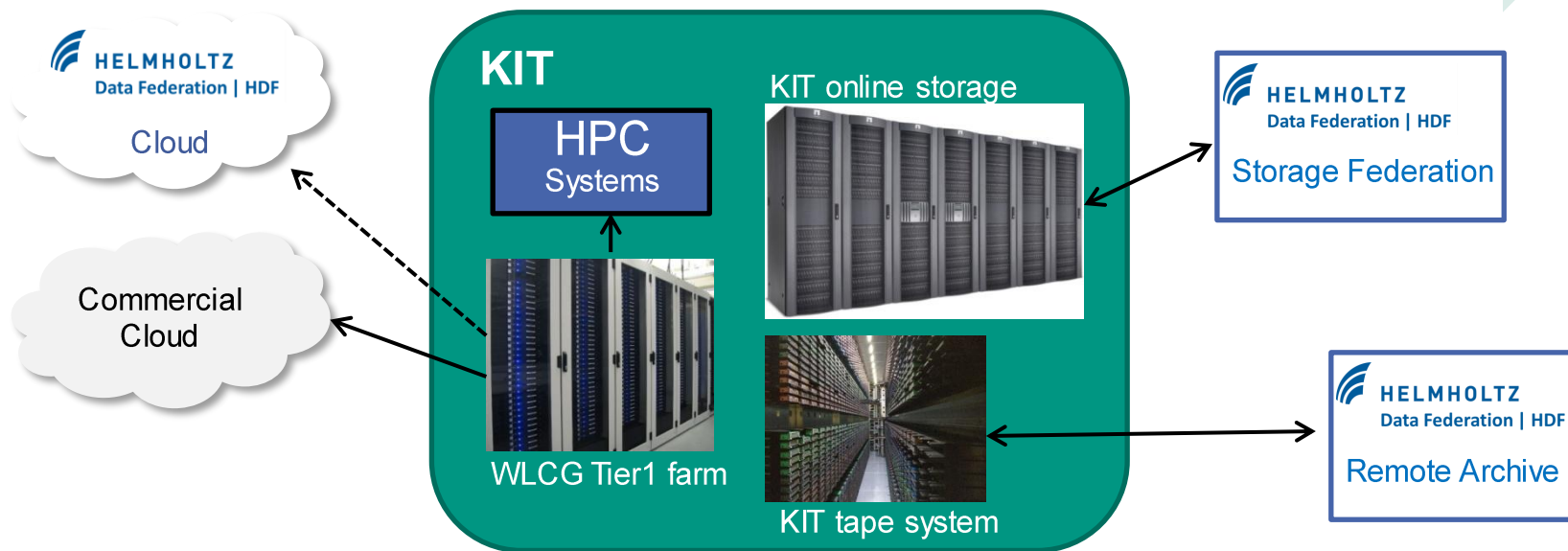 - …

- Addressing community-specific requirements
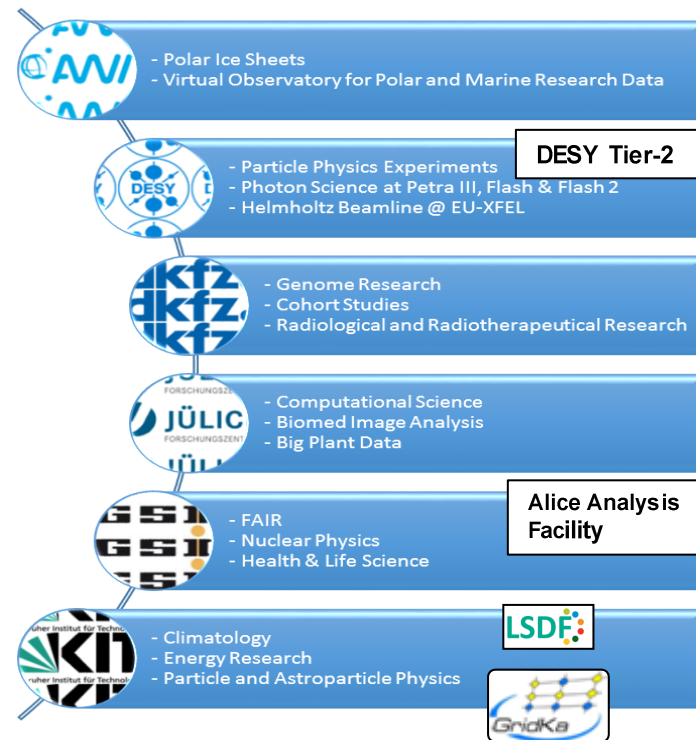
- Polar Ice Sheets
- Virtual Observatory for Polar and Marine Research Data

DESY Tier-2

- Particle Physics Experiments
- Photon Science at Petra III, Flash & Flash 2
- Helmholtz Beamline @ EU-XFEL

- Genome Research
- Cohort Studies
- Radiological and Radiotherapeutical Research

- Computational Science
- Biomed Image Analysis
- Big Plant Data

Alice Analysis Facility

- FAIR
- Nuclear Physics
- Health & Life Science

LSDF

- Climatology
- Energy Research
- Particle and Astroparticle Physics

GridKa

Integrating GridKa and LSDF

CHEP 2016, A. Heiss, A. Petzold

Steinbuch Centre for Computing

The global challenge

Steinbuch Centre for Computing

Conclusion

- Two main challenges:

  - Need to become more efficient and profit from the economy of scale
    - Integration of KIT data infrastructures

  - Rapidly changing global computing landscape
    - Many projects aiming to increase interoperability and establish common standards (EOSC, various EU funded projects)
    - Heterogeneous infrastructures
      - Own farms, clusters, clouds
      - Commercial clouds
      - HPC systems

Steinbuch Centre for Computing