

# DLCL – Matter(s)

**DESY/HTW & GSI for the communities**

**HEP, PhotonScience, Heavy Ion/Nucleon**

- currently working in community areas (mostly photon science)
  - Steve Aplin
  - Uwe Ensslin
  - Marco Strutz
  - Birgit Lewendel
  - Jürgen Starek
  - ... dCache Development Team ...

# DLCL activities @DESY (condensed)

## > HEP

- NAF(v2.0) account migration portal

## > PhotonScience

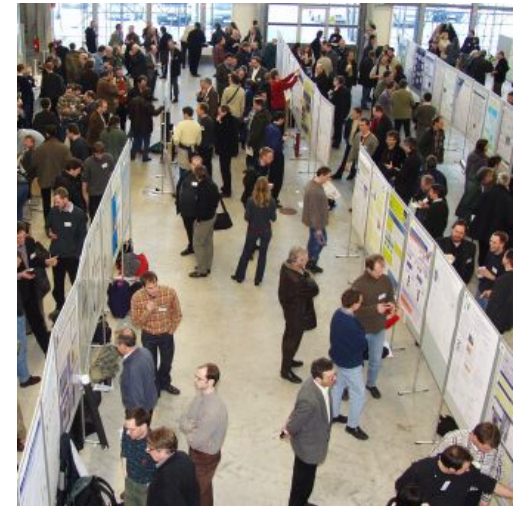
- Petra3 Beamline11 Data Recording (from detector to archive thru Analysis Storage)
- dataXpress – ‘**Processing the deluge of data from XFEL**’ CFEL + XFEL + DESY/IT
- NUMA-aware Resource Management & App development (HTW)
- ICAT Evaluation & Installation
- data management & access portal
- small file management (in dCache/Portal) for optimized Tape access
- identity management at local beamlines (beamline scientist)
- fine grained ACLs in dCache (NFSv4.1)
- NeXus (HDF5) support – i.e. metadata management and generation

## > both

- specification of generic archive service (i.e. DFG compliant)
- setup and running test environment (3 nodes/128 Cores/768GB mem/10GE)

# DLCL communications

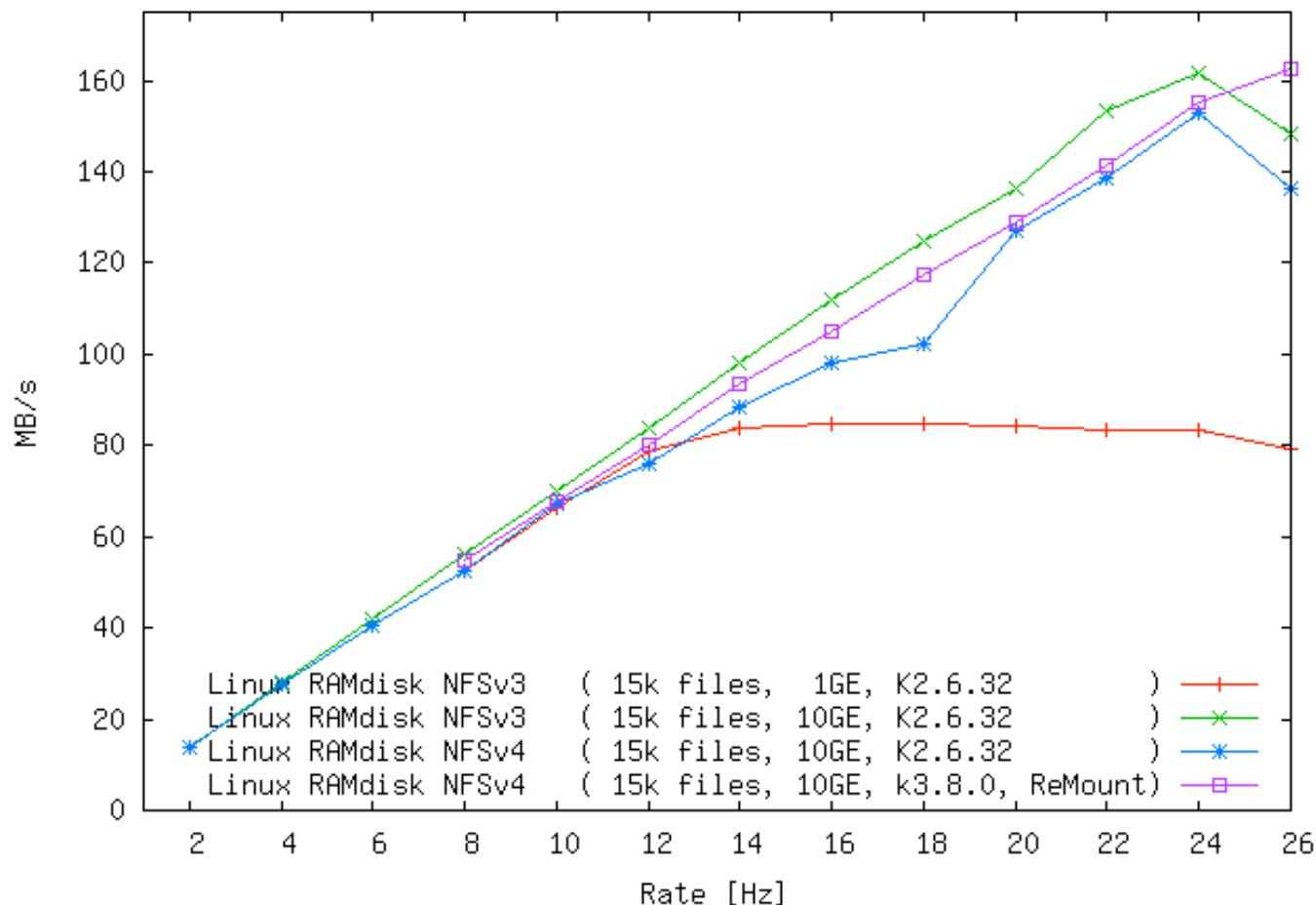
- regular meeting with local HEP, CFEL, XFEL, Petra3, Flash groups
  - in addition special meetings for detailed technical issues
- participation at HeasyLab User Meeting (January 2013)
  - more than 800 participants from worldwide
  - Posters (with LSDMA Logo !) shown
- participation in PanData/CRISP meeting (January 2013)
  - Identity management in PNI science – Umbrella





# detector to first storage @ Beamline11

NFS thruput with Pilatus (7MB) files



The applicable NFS versions on real (shared) networks. Others also include different storage systems and other protocols (i.e. gpfs).

Saturation effects below line speed.

Effects at the end unclear yet.

Serial imaging experiments at FEL facilities generate large volumes of data: sample is introduced into the FEL beam in either a liquid or aerosol jet, and intersects with X-ray pulses from the FEL at random. The detector is read out after each FEL pulse, and the diffraction pattern analysed to determine presence of sample in the beam.

**Our research group alone has collected and analysed over 1 petabyte of data from LCLS in the three year period from December 2009 to December 2012.**

LCLS 2009-2012  
(3 years)

XFEL = LCLS  $\times 10$ ?  
(120 Hz  $\rightarrow$  1200 Hz\*)

Data analysed:  
1002 TB

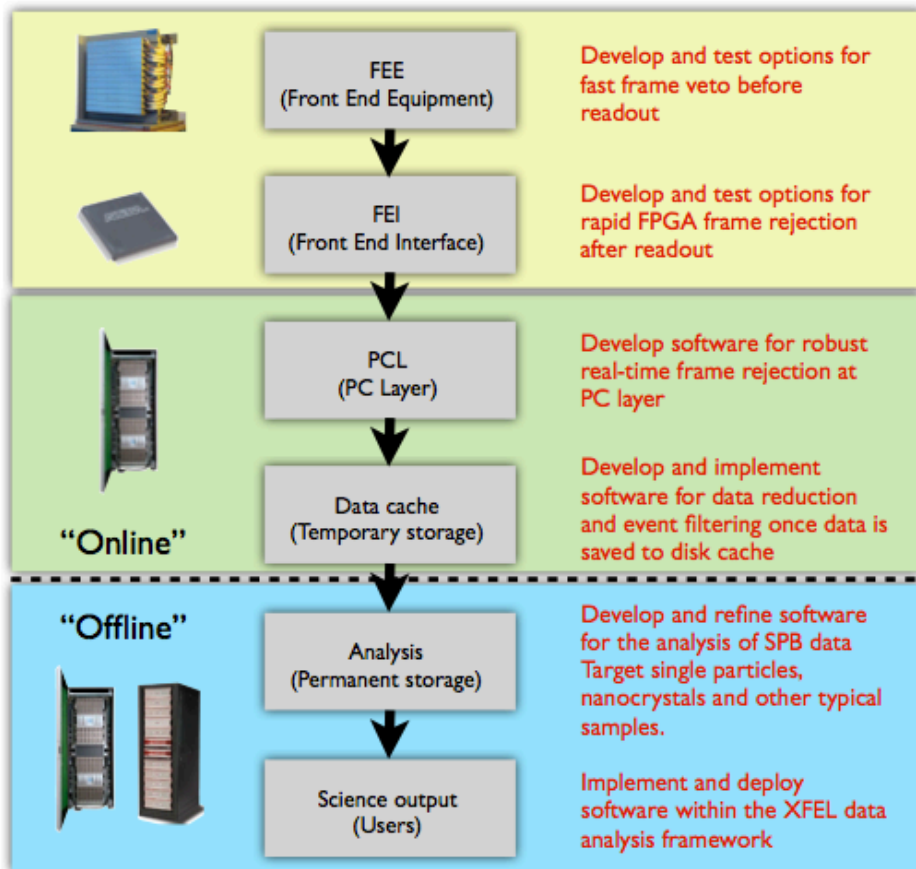
Data analysed:  
10 PB

Data taken home:  
676 TB

Data taken home:  
6PB

Reduced data:  
80 TB

Reduced data:  
1 PB



# Use Cases as of today – more to come (headline)

- > DFG compliant archive (tape based today) – bit stream preservation
- > NUMA optimizations (resources & parallel apps)
- > high speed data recording for PS beamlines
- > NeXus (HDF5) support
- > X509 auth account migration
- > System/Service accounting
- > multiprotocol access (for different profiles/requirements)
- > federated storage systems – world fs
- > small files (<10MB) – but high numbers (tape usage optimization)
- > fine grained ACLs (more than NFSv4 ACLs)

# Das Helmholtz-Project „Large Scale Data and Analysis“ (LSDMA)

Struktur der Materie/FAIR DLCL

Kilian Schwarz, GSI



## GSI computing heute

ALICE T2/T3

HADES

~ 14000 cores,

~ 5.5 PB lustre

~ 9 PB archive capacity

## FAIR computing 2018

CBM

PANDA

NuSTAR

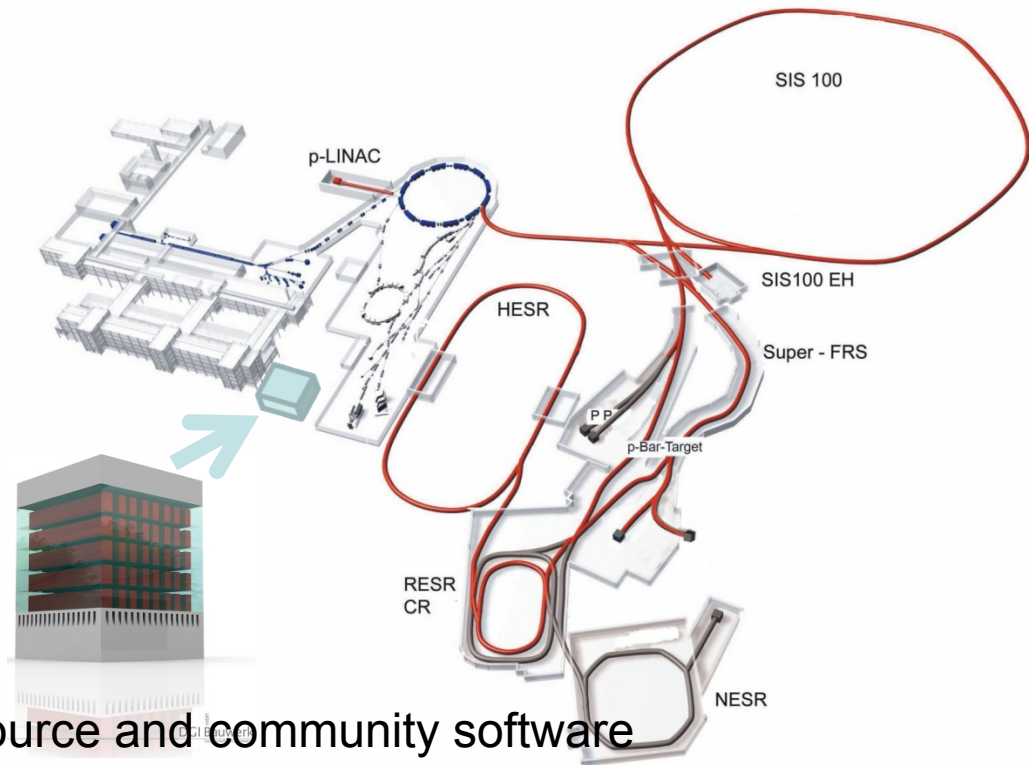
APPA

LQCD

300000 cores

40 PB disk

40 PB archive



open source and community software  
budget commodity hardware  
support different communities  
scarce manpower



For an overview with links to the experiment web pages see: <http://www.fair-center.eu/public/experiment-program.html>

FAIR Forschungsprogramm. 14 Experimente bilden die vier wissenschaftlichen Säulen:

## 1. APPA – Atomic, Plasma Physics and Applications

- **BIOMAT** – **BIO**logy and **MAT**erial science
- **FLAIR** – **F**acility for **L**ow-energy **A**ntiproton and **H**heavy **I**on **R**esearch
- **HEDgeHOB/WDM** – Plasma physics experimental stations
- **SPARC** – **S**tored **P**article **A**tomical **R**esearch **C**ollaboration

## 2. CBM/HADES – Compressed Baryonic Matter

## 3. NuSTAR – Nuclear Structure, Astrophysics and Reactions

- **R3B** – **R**eactions with **R**elativistic **R**adioactive **B**eams
- **DESPEC** – **D**ecay **S**PECTroscopy
- **HISPEC** – **H**igh resolution **S**PECTroscopy
- **MATS** – **P**recision **M**easurements of very short-lived nuclei with **A**dvanced **T**rapping **S**ystem
- **LASPEC** – **L**ASer **S**PECTroscopy of short-lived nuclei
- **ILIMA** – **I**someric beams **L**ifetimes and **M**Asses

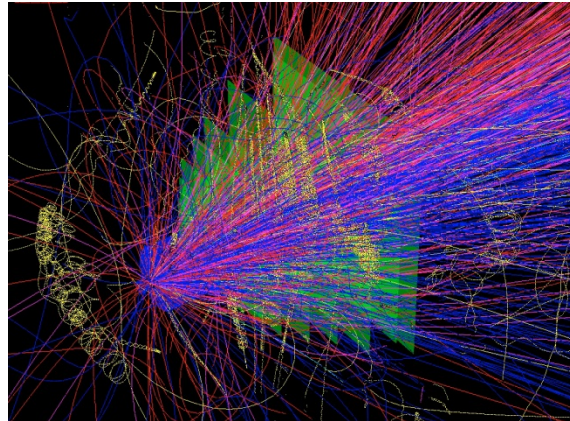
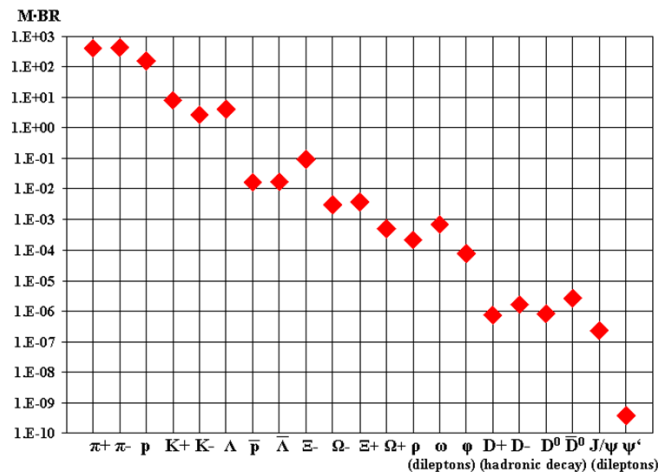
## 4. PANDA – AntiProton Annihilation in Darmstadt

FAIR hat eine stark diversifizierte User Community.

Einige der Experimente sind klein – ein paar dutzend Wissenschaftler nutzen einen Aufbau für ein paar Wochen. Andere wie PANDA und CBM sind Kollaborationen von einigen hundert Wissenschaftlern, die Vollzeit für das Experiment arbeiten und den Detektor fast das ganze Jahr über verwenden.

# Zusätzliche Herausforderungen bei der Rekonstruktion

Fast online event reconstruction, no hardware trigger:  
time to reconstruct 1-10ms &  $10^7$  collisions/s  $\rightarrow 10^4 - 10^5$  cores



Open charm:

$D^0$ ,  $c\tau=127 \mu\text{m}$

$K^-$

$\pi^+$

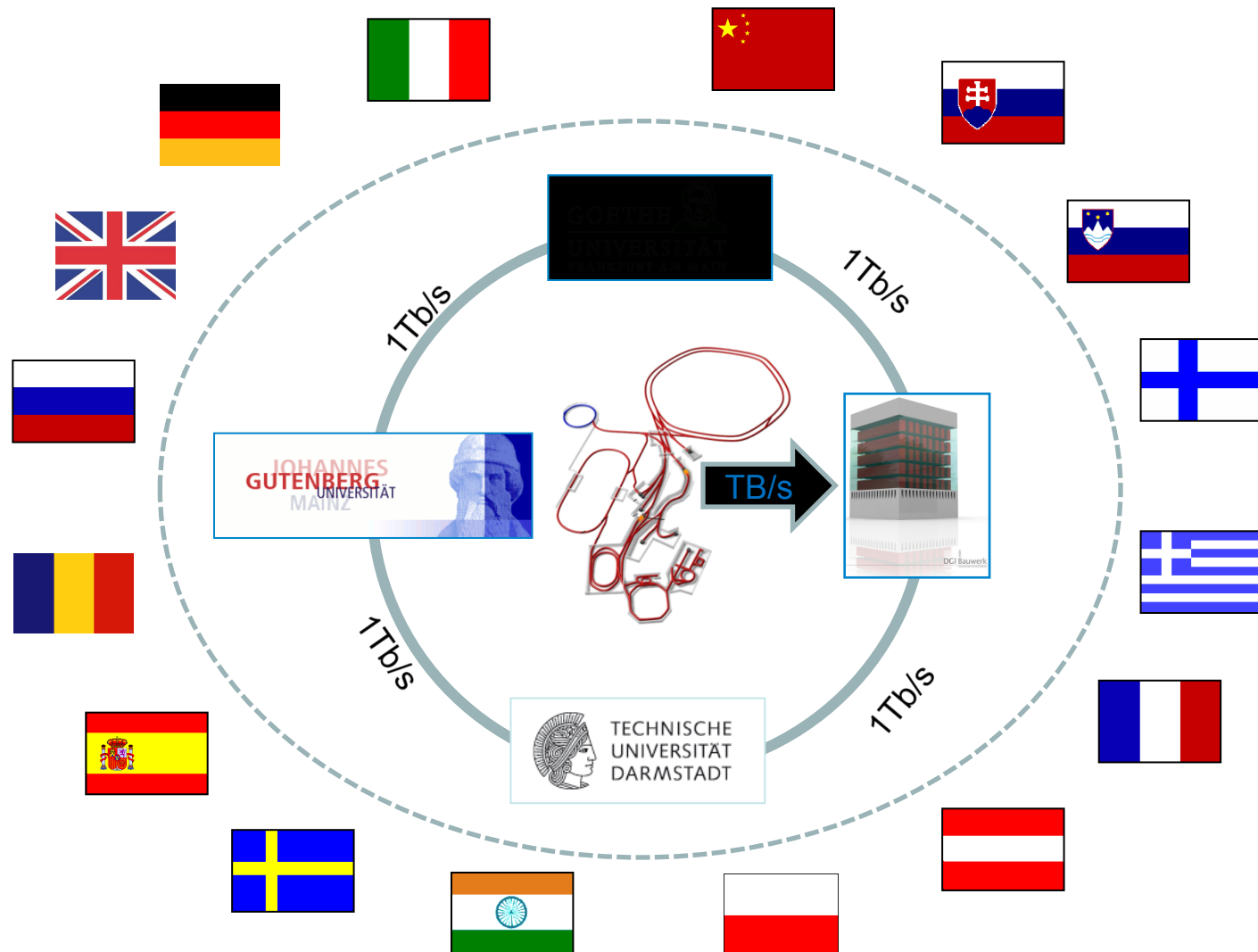
Typical signal  
multiplicities:  $O(10^{-6})$

No „easy“ trigger  
signatures

- Datennahme für CBM und PANDA bedarf einer Detektorauslese ohne Hardware-Trigger, also ausschließlich Online-Ereignis-Selektion
- Die erste Ebene des Systems besteht aus einer Kombination von spezialisierten Verarbeitungselementen wie GPUs oder FPGAs in Verbindung mit COTS Computern und einem Hochgeschwindigkeitsnetzwerk
- Die Hauptbausteine der Infrastruktur sind:
  - ein neues Rechenzentrum für Online-Computing, Speichersysteme und Offline-Systeme, eng verbunden mit
  - einem Metropolitan Area Network, welches die Computer-Einrichtungen der umgebenden Universitäten und Forschungszentren verbindet, eingebettet in
  - eine internationale Grid/Cloud-Infrastruktur



# FAIR Computing: T0/T1 MAN & Grid/Cloud



## FAIR DLCL: Allgemeine Ziele

- Paralleles und verteiltes Computing
  - Triggerloses „online“-System
  - Grid/Cloud-Infrastruktur
- Archivierung
  - Langzeitdatenarchivierung
  - Metadatenkatalog und Analyse
- Metropolitan Area Systeme
  - Authentifizierung
- Weltweite Föderationen
  - Global verteiltes Filesystem

## FAIR DLCL: Spezielle Ziele

- Paralleles und verteiltes Computing
  - Auslagern von compute-Jobs an Clouds
    - Bauen entsprechender Schnittstellen zu existierenden Umgebungen
  - Portieren notwendiger Algorithmen auf GPUs
- Archivierung
  - Erarbeitung von Archivierungskonzepten unter Einbeziehung von xrootd und gStore
- Metropolitan Area Systeme
  - Einbindung in Grid/Cloud-Infrastruktur
  - Verteilte Authentifizierung

## FAIR DLCL: Spezielle Ziele (2)

- Weltweite Föderationen
  - Optimierung der Datenspeicherung
    - Hot versus Cold Data
    - Korrupte und unvollständige Datensätze
    - Parallel-Speicherung
    - 3rd party copy
    - Global verteiltes Filesystem (Xrootd)
    - Schnittstellen zwischen Xrootd und Lustre

- Zwischen den verschiedenen Arbeitsbereichen existiert ein deutlicher Überlapp und eine klare Trennung ist schwierig.
- Es sind auch deutliche Synergien mit den DSIT-Schwerpunktthemen zu finden
- Auch mit anderen 3rd-Party-Projekten (CRISP) besteht Überlapp

- Identifizierte Schwerpunktthemen
  - Federated Identity Management
  - Federated Data Access
  - Meta-Daten
  - Archivdienst
  - Monitoring
  - Data Life Cycle
  - Verteiltes Computing

- Die GSI engagiert sich in den DSIT Workpackages 1,2 und 5.

Im Rahmen von WP1 (Federated Identity Management) finden momentan Sondierungen im Hinblick auf die Anforderungen der großen FAIR-Kollaborationen sowie CRISP WP16, an dem sich die GSI ebenfalls beteiligt, statt.

Im Rahmen von DSIT WP2 (Federated Data Access) wird an der Absicherung von Lustre mittels Kerberos, um dieses für einen abgesicherten WAN-Zugriff zu ertüchtigen, gearbeitet. Ein Modul um GID und UID Informationen zwischen Sites für die Benutzer transparent transformieren zu können, ist in Arbeit.

Ein prototypischer Remote-Zugriff über eine breitbandige WAN-Verbindung zwischen GSI und LOEWE-CSC Uni Frankfurt per Lustre wurde aufgebaut.

Performance-Monitoring und -Optimierung dieses Zugriffs werden im Rahmen von WP5 ( Monitoring) stattfinden.

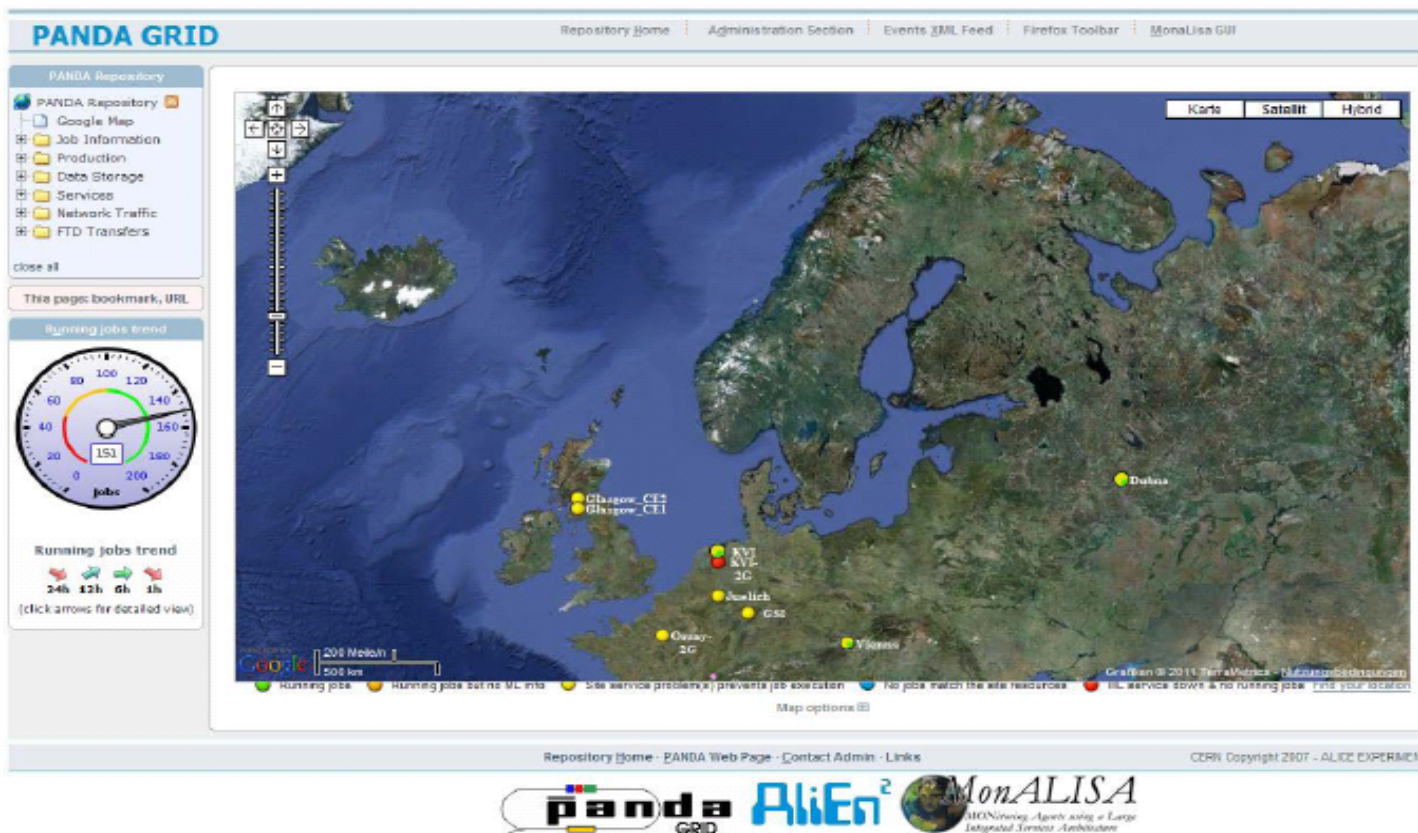
- LSDMA Stellen ausschreiben (2 für FAIR DLCL)
- Diskussion (Bedarfsanalyse) mit FAIR Experimenten und ALICE
- Aufbau einer e-science - Umgebung, zunächst für PANDA and CBM, basierend auf den Erfahrungen mit ALICE (AliEn/xrootd/...)
- Einbeziehung kleinerer FAIR-Experimente
- Weiterentwicklung der existierenden e-science – Umgebung, auch in enger Zusammenarbeit mit DSIT und ALICE



- Aktuell arbeiten 0,5 FTE für FAIR DLCL
- Bedarfsanalyse mit NuStar ist erfolgt
- Entwicklung von modular darstellbarer Detektorgeometrie für FairRoot
  - wichtig für NuStar-Datenanalyse
- PandaGrid ist im Produktionsbetrieb
  - Basis für weitere Fair-Experimente ?
- LSDMA-Übersichtsvortrag bei DPG-Frühjahrstagung in Dresden 2013

## Grid & Cloud for FAIR

PandaGrid, Cloudification of LOEWE CSC, Ffm Cloud, ...



- Es gibt viele Herausforderungen in Zusammenhang mit großen Datenmengen
- LSDMA ist ein Projekt des Helmholtz-Verbundes, welches den gesamten Lebenszyklus der Daten unterstützt und dabei einen dualen Ansatz hat (community-spezifisch und generisch)
- FAIR ist eine wichtige Community im Forschungsumfeld “Struktur der Materie”