e-Infrastructures & Big Data Handling

as a cross topic challenge within CREMLIN

Volker Guelzow
DESY
Moscow, Oct 7th, 2015

We all know:

No computing

means

> no experiments

> no science

> Therefore: „computing" has to be included from the beginning
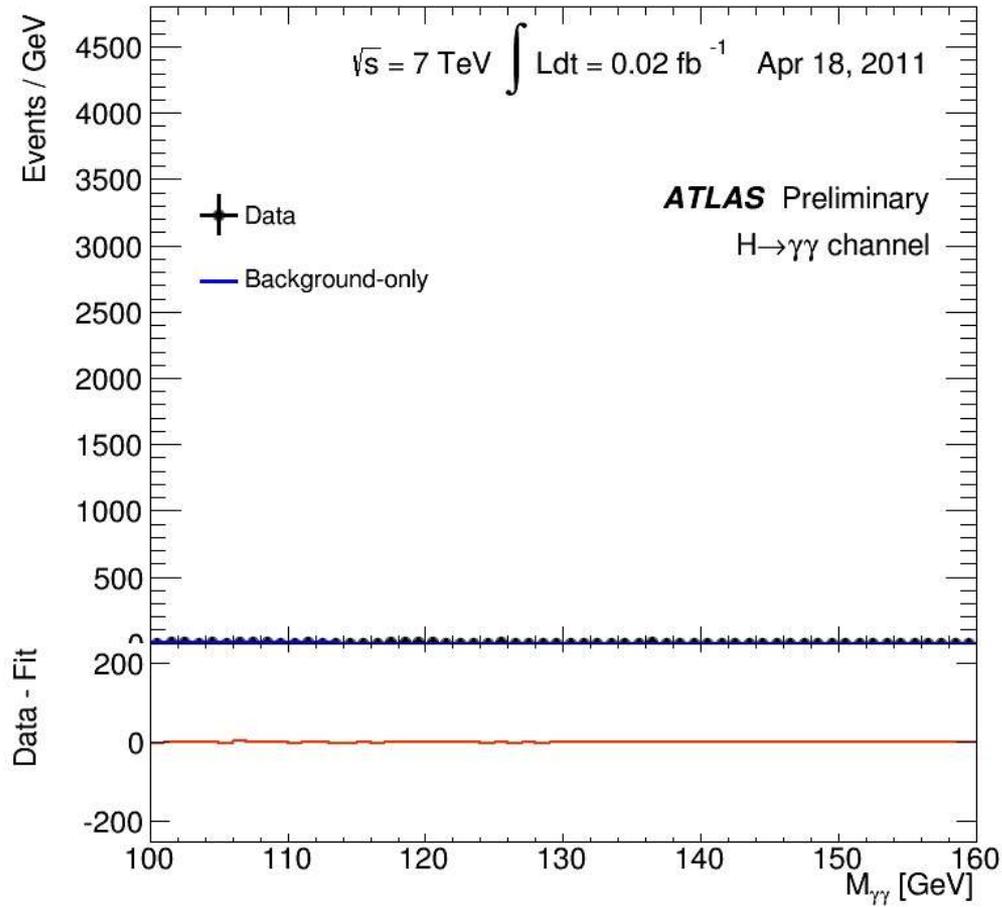
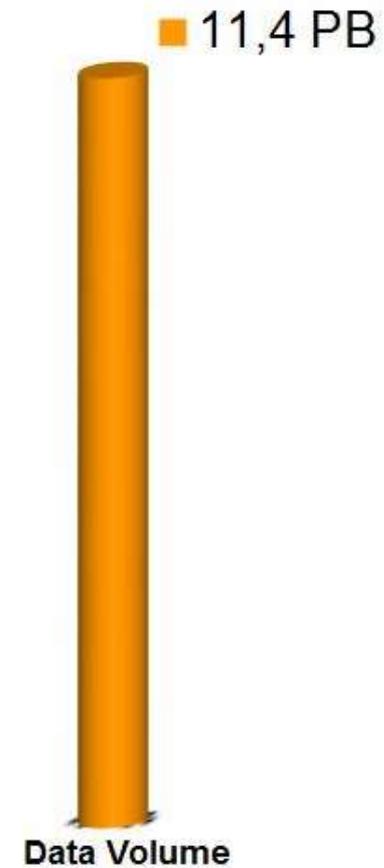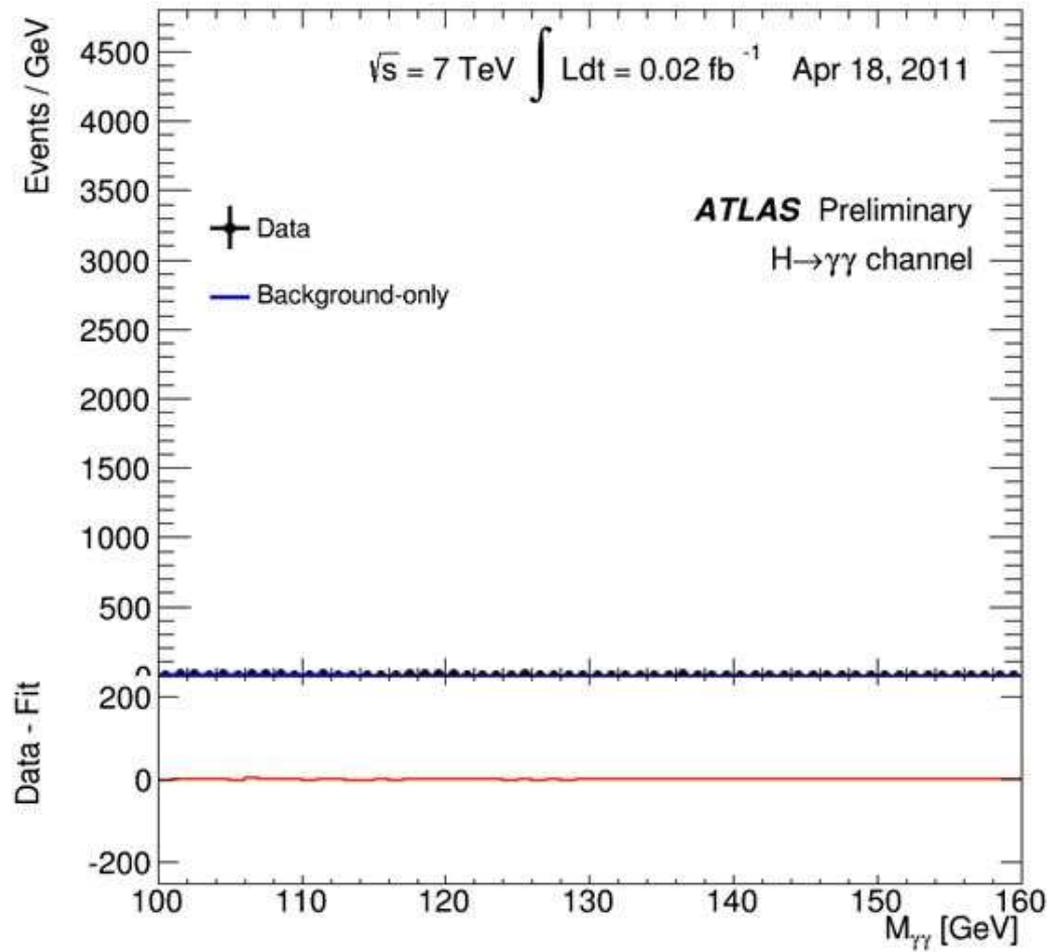> Computing is a „cross workpackages" topic in CREMLIN

… but many common issues

Higgs Discovery

Higgs Discovery

Limnephilus flacivornis Head & Thorax

Courtesy:
Dr. F. Beckmann

The ICT Eco-System for large facilities

Data&Metadata Management Software

Simulations

Detector

Outside

Local Cache

Central Storage

Online Archive

Data Analysis

- Grid
- AF
- HPC
- Farm
- Cloud
- Visuali zation

Local Researcher

Remote Researcher

Data Policies

User Management/AAI

Technical Infrastructure

networks
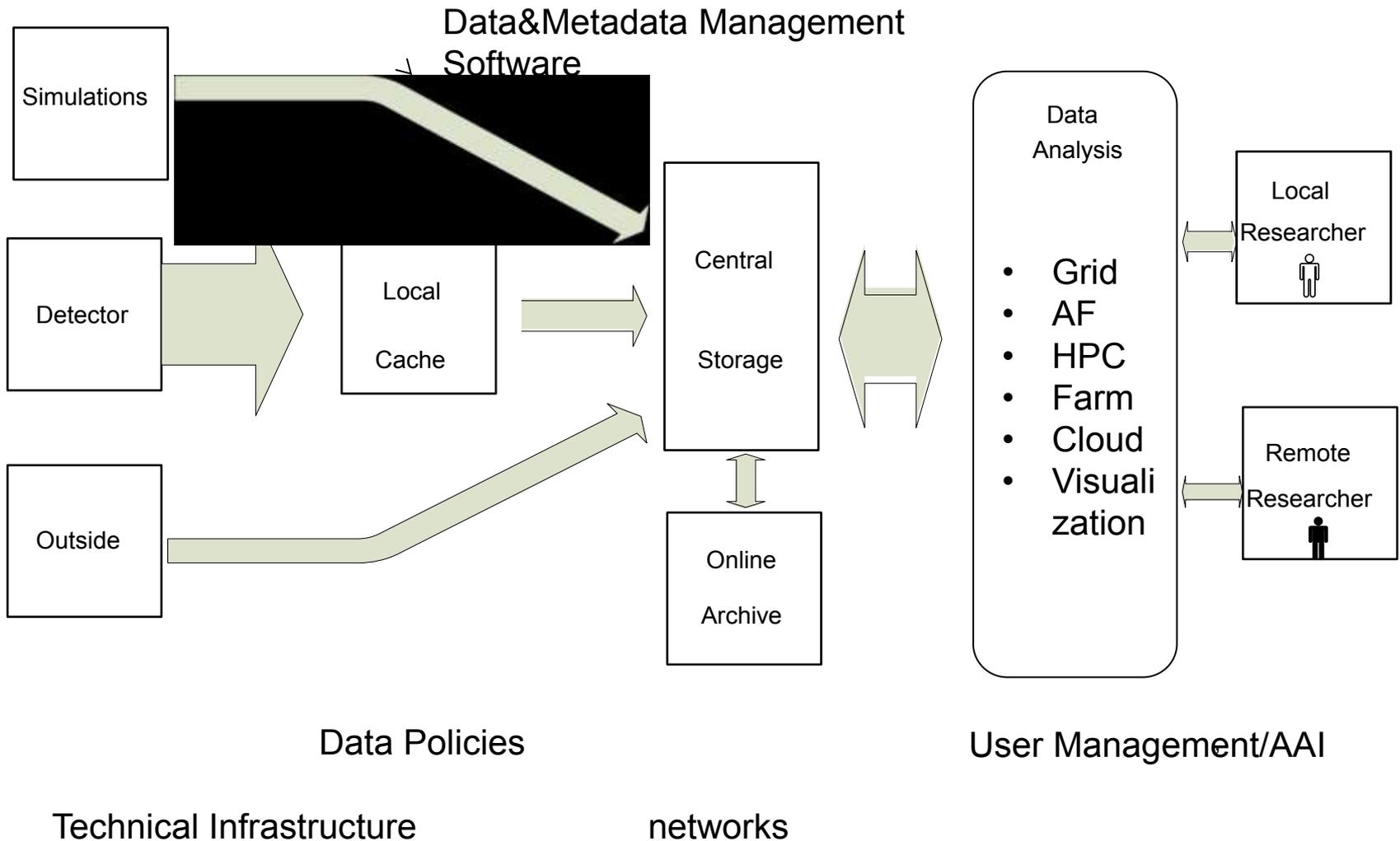
New Challenges

Type	Frame size	Frame rate	Peak rate	Avail.
Pilatus 6M	2463 x 2527 x 4	25 Hz	4.6Gb/s	Now
AGIPD (Module)	128 x 512 x 2 x 352 x 14bit	4.5 MHz (10 Hz)	6.1 Gb/s	2015
Eiger	1k x 1k x 2	2 kHz	30 Gb/s	now
Lambda	3 x 1536 x 512 x 2	2 kHz	60 Gb/s	now
Percival (1S)	4k x 4k x 2	120 Hz	60 Gb/s	2015
Percival (4S)	8k x 8k x 2	120 Hz	240 Gb/s	Late 2015
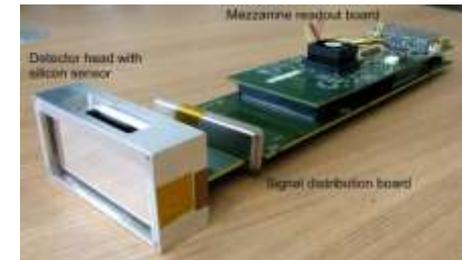
The Speed Project for Beamlines:

Goal:

- Very fast data management for Photon Science
- Development at Petra III, prototype for EU-XFeL
- Cooperation DESY/IBM
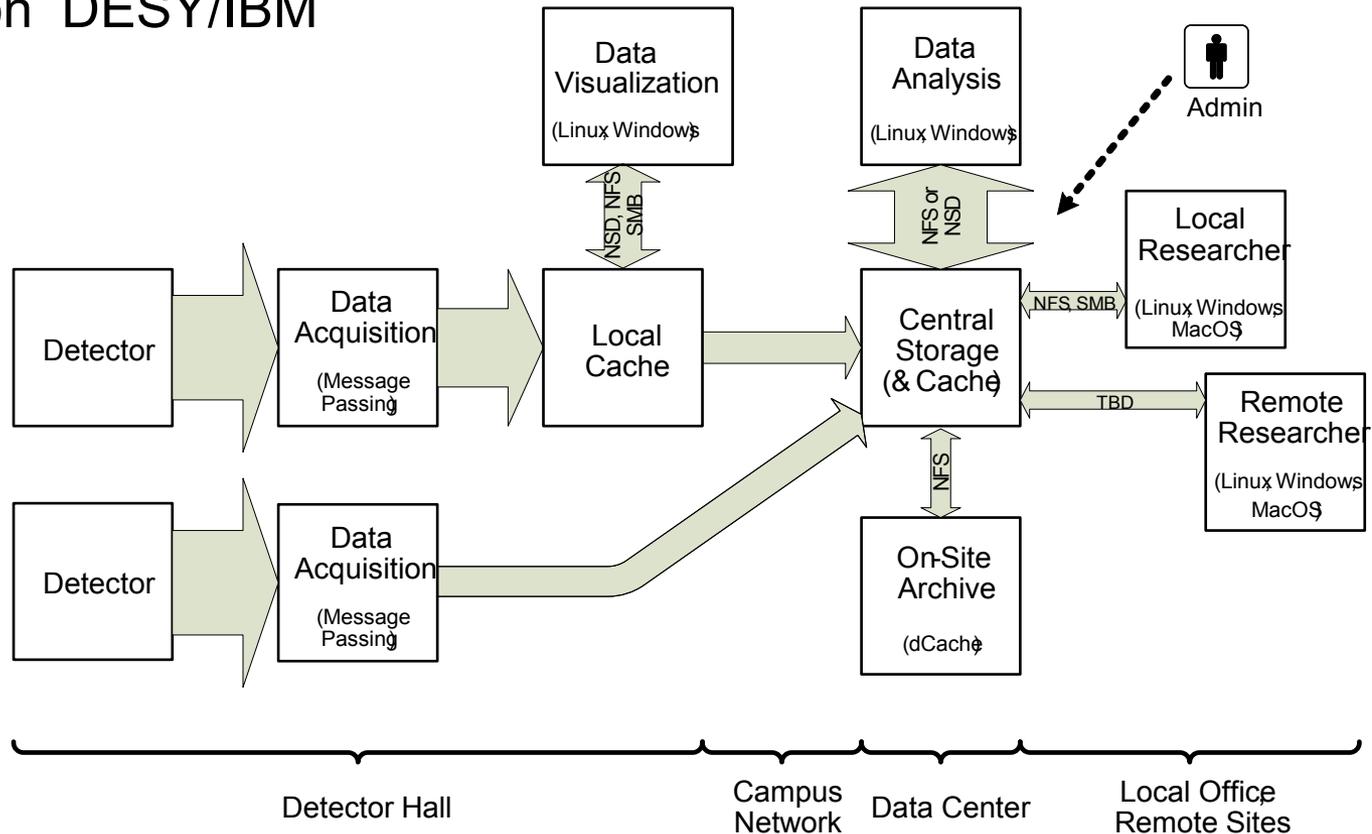
Important elements for ICT for megascience infrastructures

> Computing is fairly easy, the software is the key issue

> Develop high sophisticated „Big Data" management solutions

  - … fitting on site experiments and off site experiments

  - Provide sufficient hardware resources (capacity and speed)

  - Sufficient bandwidth

  - Think about data formats (eg NEXUS in Photon Science)

  - Set up a Authorization and Authentication Infrastructure (AAI)

  - Easy access (like Eduroam,…)

  - Portals to access the data from remote

  - Datapolicies (owner, store how long, who has access, ……)

  - Long term data preservation

  - Develope solutions in cooperation with other Lab's

> New „kids on the block" like clouds….

> Don't forget about the energy

dCache Big Data Cloud

Intensity frontier

Supercomputing
for Human Brain

LOFAR antenna
Huge amounts of data

Cloud User

WebDAV
HTTP(S)

Globus Online
Cloud

dCache

Disks

Tape Storage

Mounted POSIX FS
(NFSv4.1, pNFS)

Computer farm

European XFEL

X-FEL
(Free Electron Lasers)
Fast Ingest
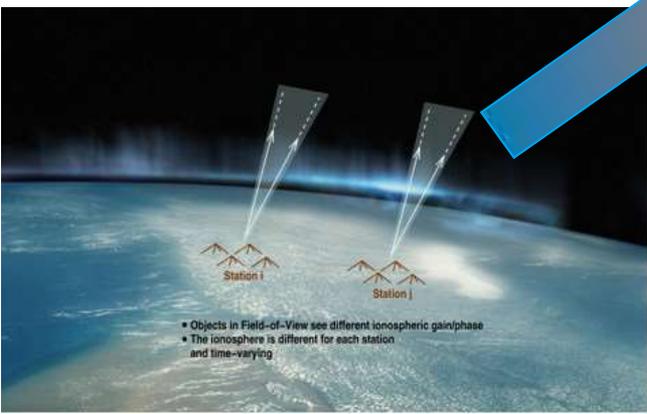
WEB 2.0

ownCloud

Unlimited hierarchical
Storage Space
NFS 4.1
CDMI

dCache
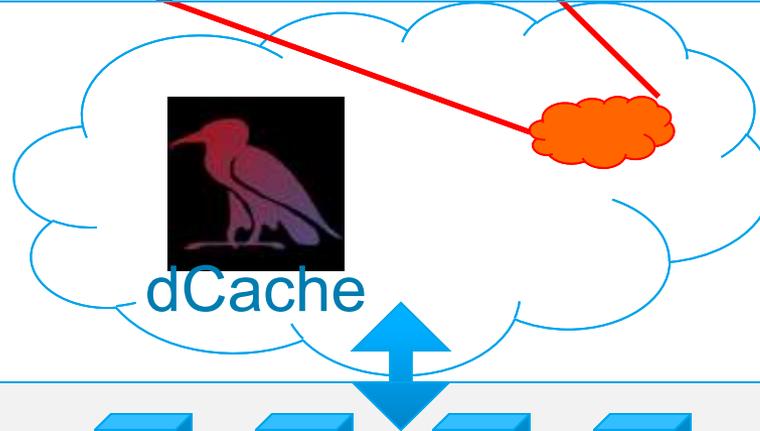
SSD'S

Spinning Disks

Tape, Blue-ray …

DESY

GÉANT connectivity as at January 2014.

At the Heart of Global Research and Education Networking

GÉANT and partner networks enabling user collaboration across the globe
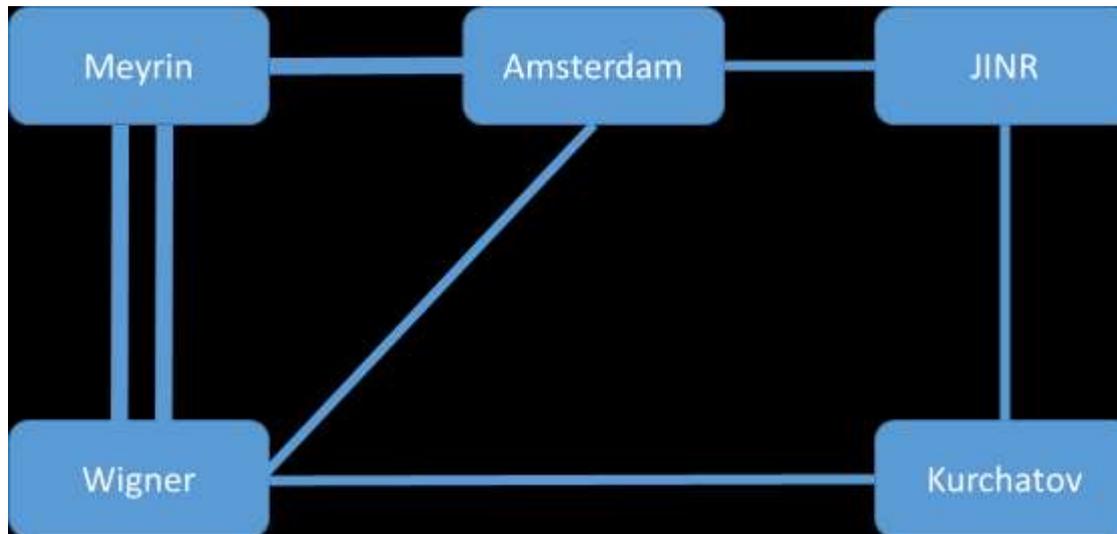
September 2014

connect • communicate • collaborate
GÉANT is co-funded by the European Union within its 7th R&D Framework Programme.

Connectivity

> Existing links for LHC:

> JINR has a 10G link to Amsterdam (and than via 100 G to Cern)

> Kurchatov has 10 G to the Wigner Centre

> Tier 2 connectivity goes via NorduNet or ESNET but not Geant

> Connectivity is essential for megascience facilities (10-100 Gb/s)

Conclusion

> There is no „one fits all" solution

> if possible, use standard data formats

> All participating centres do have a outstanding knowledge in ICT, let's put things together

> common data management solutions can form a basis for big data management

> A proper user interface is mandatory

> Connectivity to the world is a key factor

> Well supported by the EC