

FACILITY TOPIC

IDAF
M&T LK II Facility@DESY
Christian Voß



Motivation



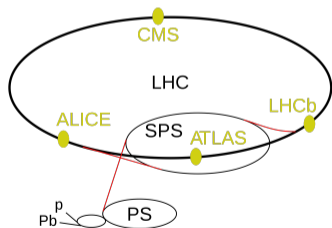
Data Driven Sciences – LHC Experiments

- Science in *Matter* is largely data driven → need facilities to analyse the data
- Existing Facility: MU Tier-2 Centre → one of the largest sites within WLCG

Grid Cluster: ■ Usage driven by LHC experiments



- Tier-2 Centre for ATLAS, CMS, LHCb; raw data centre for Belle II (from 2021)
- Large scale analysis work with long turn around cycles

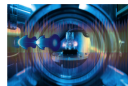


Experimental Data

National Analysis Facility

- Usage driven by individual scientists
- Part of large collaborations
- Full access to the Tier-2 storage
- Designed for interactivity → short turnarounds
- Well established toolkits

Motivation



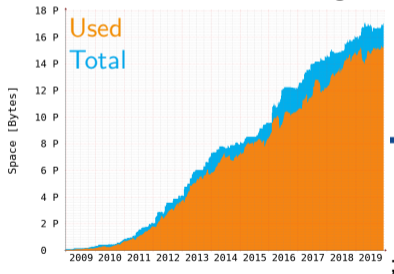
Data Driven Sciences – LHC Experiments

- Science in *Matter* is largely data driven → need facilities to analyse the data
- Existing Facility: MU Tier-2 Centre → one of the largest sites within WLCG

Grid Cluster: ■ Usage driven by LHC experiments



- Tier-2 Centre for ATLAS,CMS, LHC*b*; raw data centre for Belle II (from 2021)
- Large scale analysis work with long turn around cycles



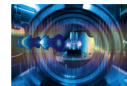
Experimental Data

National Analysis Facility

- Usage driven by individual scientists
- Part of large collaborations
- Full access to the Tier-2 storage
- Designed for interactivity → short turnarounds
- Well established toolkits

Motivation

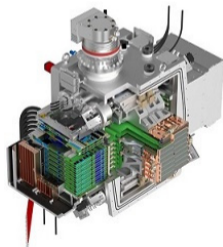
Data Driven Sciences – Photon Science Experiments



- Dedicated High Performance Computing infrastructure

Online Computing and Calibration:

- Driven by local detectors and users from national and international institutes
- Different solution for different beam-lines
- Limited automation



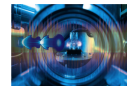
Experimental Data

User Analysis

- Usage driven by individual scientists
- Not part of large collaborations (word-wide access needed)
- Access to all the experimental data
- Large interactive Analysis
- Lack of unifying toolkits

Motivation

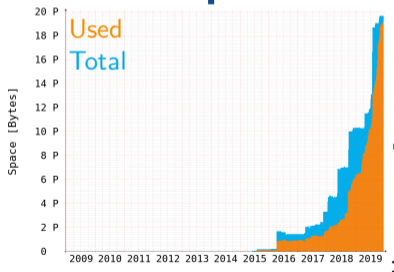
Data Driven Sciences – Photon Science Experiments



- Dedicated High Performance Computing infrastructure

Online Computing and Calibration:

- Driven by local detectors and users from national and international institutes
- Different solution for different beam-lines
- Limited automation

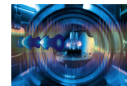


Experimental Data

User Analysis

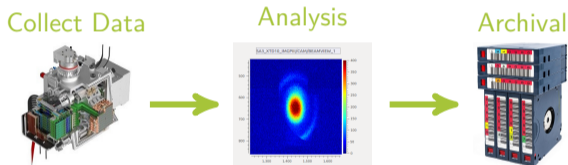
- Usage driven by individual scientists
- Not part of large collaborations (word-wide access needed)
- Access to all the experimental data
- Large interactive Analysis
- Lack of unifying toolkits

Motivation

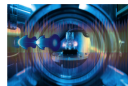


Overlap Between All Fields of Matter

- Requirements from MML experiments reach scale of LHC → necessity to improve on existing methods and ideas
- Better detectors deliver more data → data reduction becomes an issue

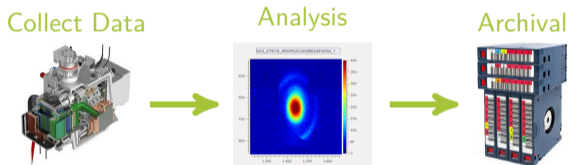


Motivation



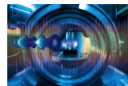
Overlap Between All Fields of Matter

- Requirements from MML experiments reach scale of LHC → necessity to improve on existing methods and ideas
- Better detectors deliver more data → data reduction becomes an issue



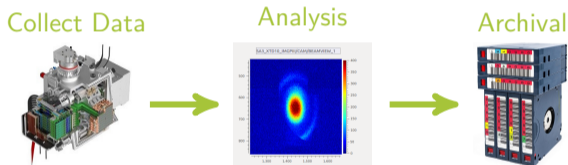
- New accelerators and detectors developed within MT become ever more complex
- Simulating these new generations requires more and more resources
- These produce large amounts of sensor data themselves → increasing demand for resources

Motivation



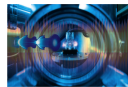
Overlap Between All Fields of Matter

- Requirements from MML experiments reach scale of LHC → necessity to improve on existing methods and ideas
- Better detectors deliver more data → data reduction becomes an issue



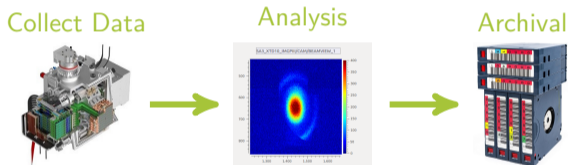
- New accelerators and detectors developed within MT become ever more complex
- Simulating these new generations requires more and more resources
- These produce large amounts of sensor data themselves → increasing demand for resources
- HL-LHC requires restructuring of computing throughout WLCG → Exascale Computing

Motivation



Overlap Between All Fields of Matter

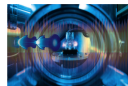
- Requirements from MML experiments reach scale of LHC → necessity to improve on existing methods and ideas
- Better detectors deliver more data → data reduction becomes an issue



- New accelerators and detectors developed within MT become ever more complex
- Simulating these new generations requires more and more resources
- These produce large amounts of sensor data themselves → increasing demand for resources
- HL-LHC requires restructuring of computing throughout WLCG → Exascale Computing

Strong incentive for interdisciplinary collaboration across *Matter*

From Tier-2 and HPC Cluster to the IDAF

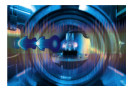


Establishment of the Interdisciplinary Data Analysis Facility

- Artificial separation determined by **community** not by **requirements**
 - (Astro-)Particle Physics community (MU) present on Grid&NAF systems
 - Machine and Photon Science (MT and MML) present on HPC system
- **Assign resources through requirements**
- HPC resources for Machine Learning (GPUs) easily accessible for all scientists
- Non-HPC applications from MML access to the NAF to free up HPC resources
- *Matter* as an interdisciplinary program suggests creation of an interdisciplinary facility
- Shared infrastructure leads to shared knowledge
 - Experience through Machine Learning seminars
 - Scientists use similar tools on same GPU hardware

Objectives for the IDAF

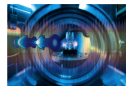
Integrating Different Communities into a Single Infrastructure



- Simply opening up resources for all communities **not sufficient**
 - Access to classic compute resources can feel clunky
 - Moving from Grid&NAF \iff HPC requires learning of different submission tools
 - Data access requires a certain technical expertise
- Infrastructure itself needs changes
 - close collaboration with MT-DMA and all other topics in *Matter*

Objectives for the IDAF

Integrating Different Communities into a Single Infrastructure



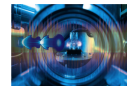
- Simply opening up resources for all communities **not sufficient**
 - Access to classic compute resources can feel clunky
 - Moving from Grid&NAF \iff HPC requires learning of different submission tools
 - Data access requires a certain technical expertise
- Infrastructure itself needs changes
 - close collaboration with MT-DMA and all other topics in *Matter*

Objectives to support all scientists within *Matter* across the IDAF

- Support efficient access with maximum flexibility to compute and storage resources
- Workflows and Tool need to be invariant when scientists switch infrastructure
- Access to data transparent and simplified
- Minimally invasive to well established workflows

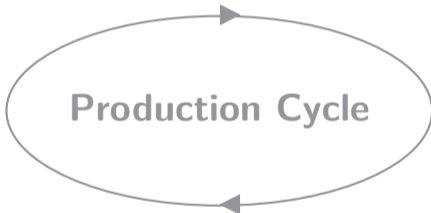
IDAF – How to Achieve the Objectives

Overlapping Expertise between LK I and LK II activities especially MT-DMA



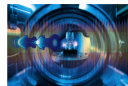
Expert Scientists and MT-DMA

User Communities
(from all *Matter* fields)



Interdisciplinary Data and Analysis Facility

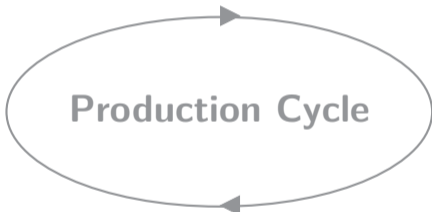
IDAF – How to Achieve the Objectives



Overlapping Expertise between LK I and LK II activities especially MT-DMA

Expert Scientists and MT-DMA

- Collect user feedback and applications

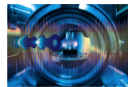


- Production application support

Interdisciplinary Data and Analysis Facility

User Communities
(from all Matter fields)

IDAF – How to Achieve the Objectives



Overlapping Expertise between LK I and LK II activities especially MT-DMA

Expert Scientists and MT-DMA

- Collect user feedback and applications

Production Cycle

- Production application support

Development Cycle

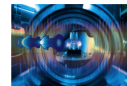
Interdisciplinary Data and Analysis Facility

User Communities
(from all Matter fields)

Emerging Technologies
(internal/external impulses)

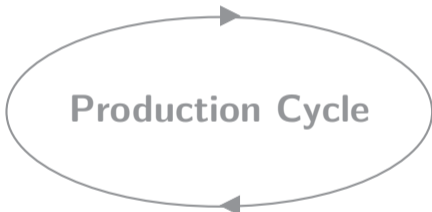
IDAF – How to Achieve the Objectives

Overlapping Expertise between LK I and LK II activities especially MT-DMA



Expert Scientists and MT-DMA

- Collect user feedback and applications
- Identify key emerging technologies



- Production application support



- Prototype on dedicated resources

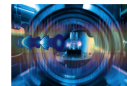
Interdisciplinary Data and Analysis Facility

User Communities
(from all Matter fields)

Emerging Technologies
(internal/external impulses)

IDAF – How to Achieve the Objectives

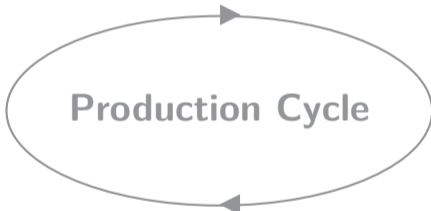
Overlapping Expertise between LK I and LK II activities especially MT-DMA



Expert Scientists and MT-DMA

- Collect user feedback and applications

- Identify key emerging technologies
- Refine prototypes into pilots



- Production application support

- Prototype on dedicated resources
- Pilot on production resources

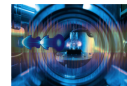
Interdisciplinary Data and Analysis Facility

User Communities
(from all Matter fields)

Emerging Technologies
(internal/external impulses)

IDAF – How to Achieve the Objectives

Overlapping Expertise between LK I and LK II activities especially MT-DMA



Expert Scientists and MT-DMA

- Collect user feedback and applications
- Improve systems with user feedback

- Identify key emerging technologies
- Refine prototypes into pilots

Production Cycle

- Production application support
- Adapt into production

Development Cycle

- Prototype on dedicated resources
- Pilot on production resources

Interdisciplinary Data and Analysis Facility

User Communities
(from all Matter fields)

Emerging Technologies
(internal/external impulses)

Competences



For Establishing the Interdisciplinary Data Analysis Facility

- Existing Infrastructure supports already user groups across all fields of *Matter*
- Tier-2 centre major contributor among all WLCG sites with regard to CPU and storage
- HPC cluster handles the online computing for PETRA-III/FLASH and EuXFEL
- HPC cluster supporting Photon Science Users at PETRA-III/FLASH and EuXFEL

Competences

For Establishing the Interdisciplinary Data Analysis Facility



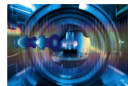
- Existing Infrastructure supports already user groups across all fields of *Matter*
- Tier-2 centre major contributor among all WLCG sites with regard to CPU and storage
- HPC cluster handles the online computing for PETRA-III/FLASH and EuXFEL
- HPC cluster supporting Photon Science Users at PETRA-III/FLASH and EuXFEL

Most Members of *Matter* are active on the existing infrastructure

Some examples: Anton Barty (CFEL), Frank Gaede (DESY FH Division/FLC), Yves Kemp (MT-DMA ST-1 Spokesperson, DESY IT Division), Jens Osterhoff (MT-ARD Spokesperson), Christian Voß (IDAF Spokesperson, DESY IT Division), Tim Wilksen (MT-DMA ST-3 Speaker, DESY M Division)

Conclusions

on the Interdisciplinary Data Analysis Facility



- Existing infrastructure delivers resources to allow scientist to harvest data for excellent scientific results
- Existing infrastructure is a valid starting point
- The program *Matter* benefits from a overarching shared computing infrastructure
- Exploit close connection to scientists in *Matter*
- Even deeper connection with the new topic MT-DMA