



Interdisciplinary

Data and

Analysis

Facility.

Strategy



Photon science **MML** Accelerator R&D **MT**
(Astro-)particle physics **MU**



Data centric concepts and implementation



Designed for the diversity of analyses



Supporting the complete data lifecycle

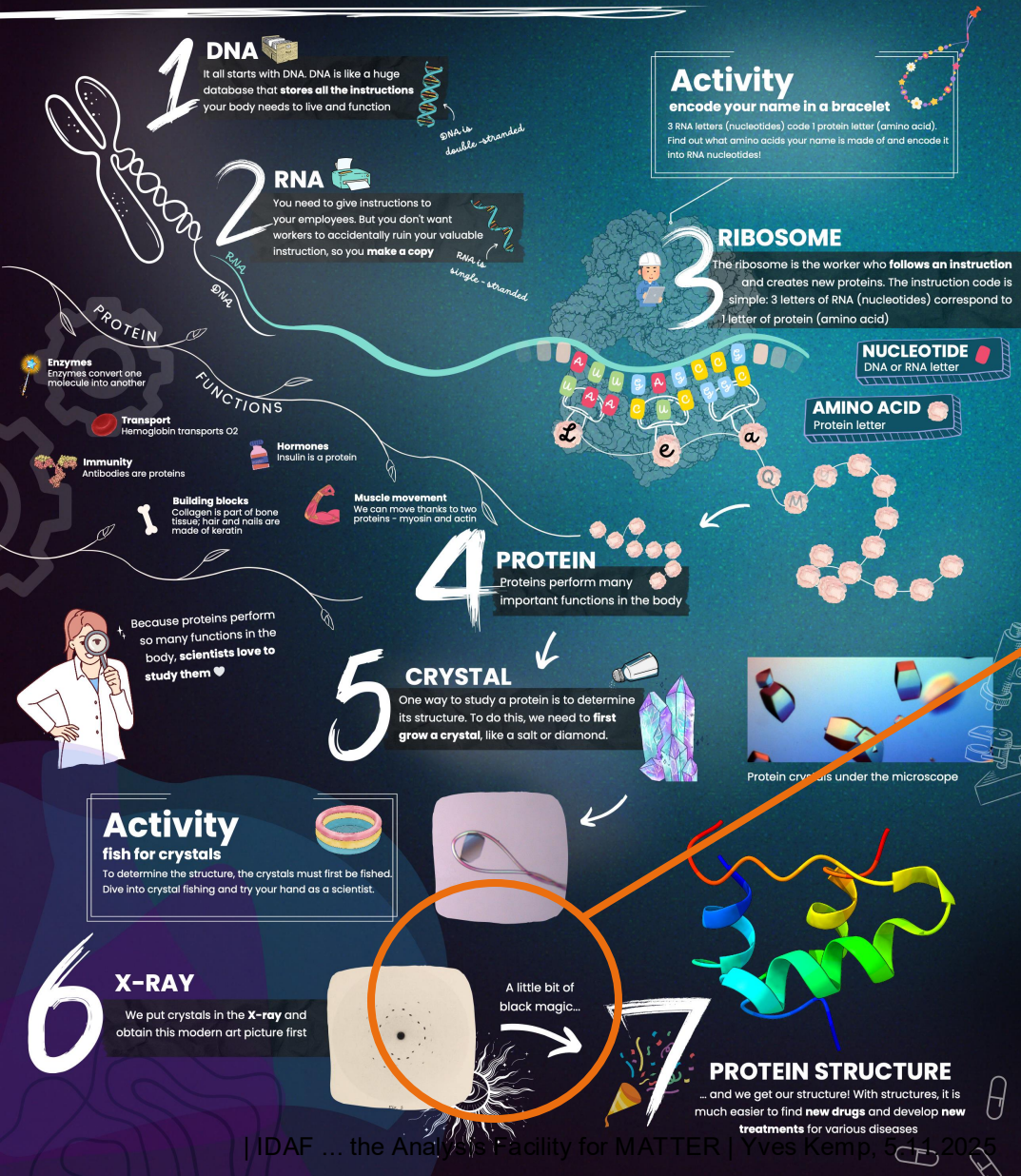
Yves Kemp, Christian Voß, Philipp Neumann / DESY-IT
GSI, 5.2.2025

HELMHOLTZ

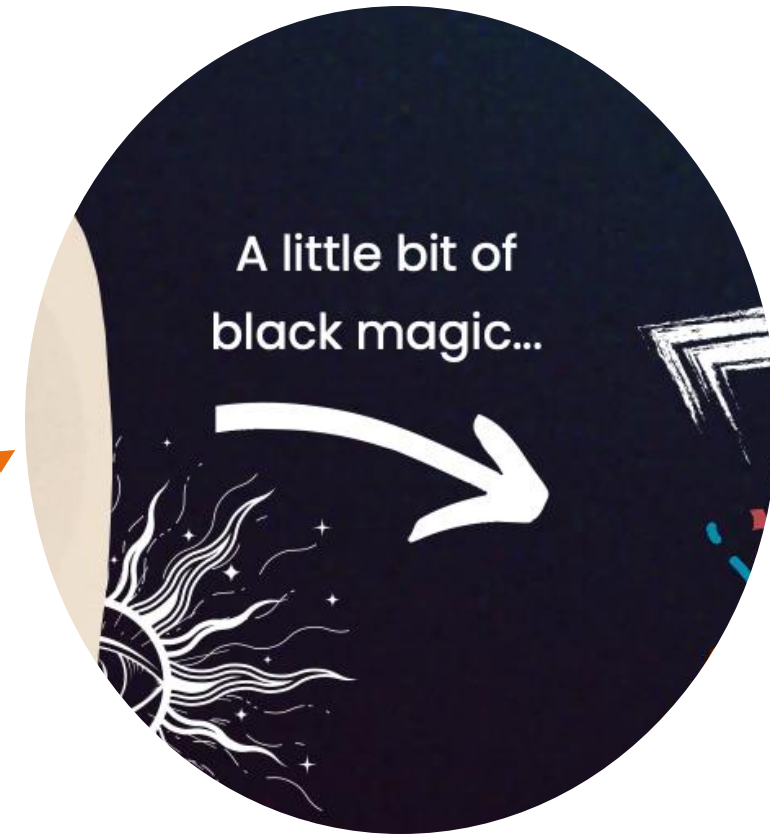


Dive into science

or the shared secrets of salt, diamond, and protein structure



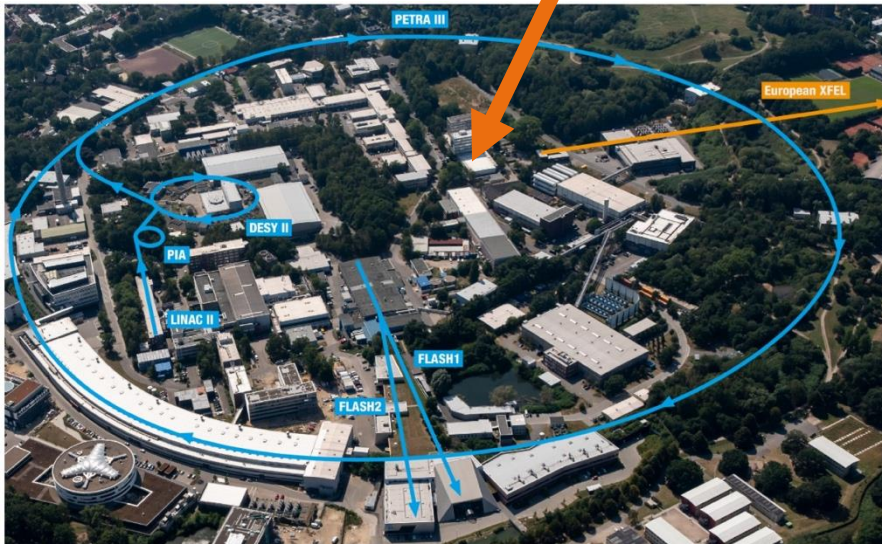
courtesy: Tatiana Kupriianova, Ph.D. student, Garcia Alai Team, EMBL
Public poster, 1.6.2024 (Science City Day)



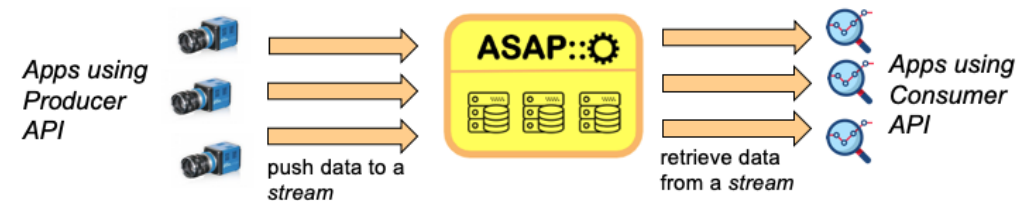
IDAF supports photon science

Integrated into data taking of:

- PETRA III
- EuXFEL
- FLASH



Enables high demand data taking:

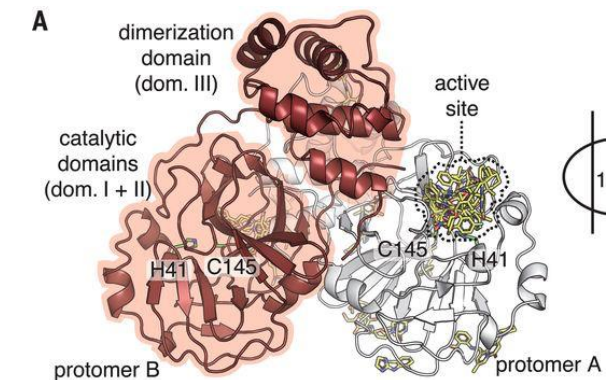


Enables data analysis: e.g.:

“X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease”

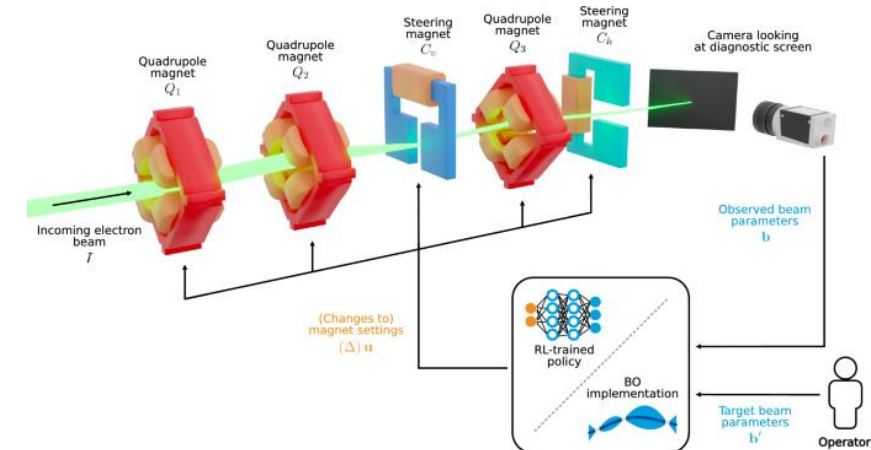
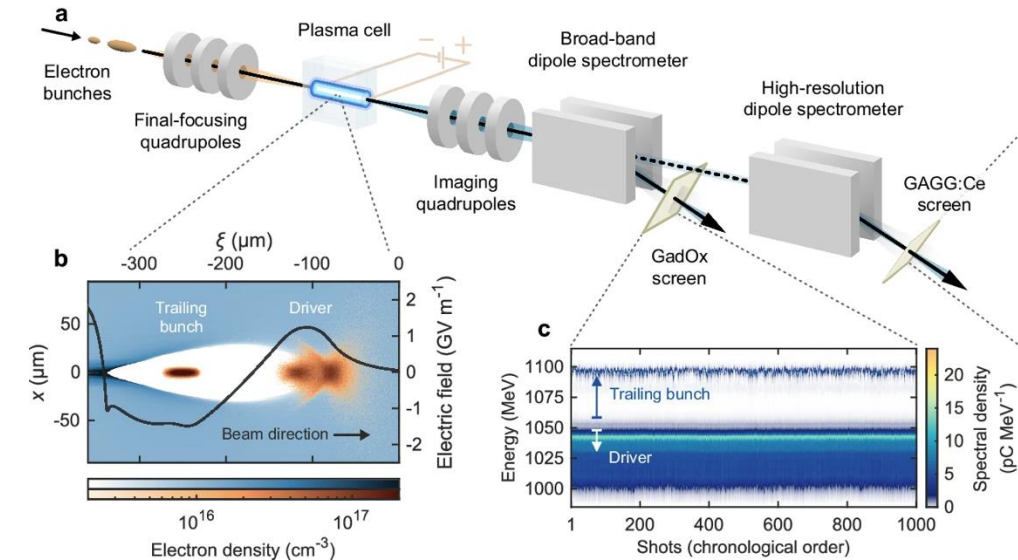
Sebastian Günther et al.

Science **372**(6542):642-646 (2021)

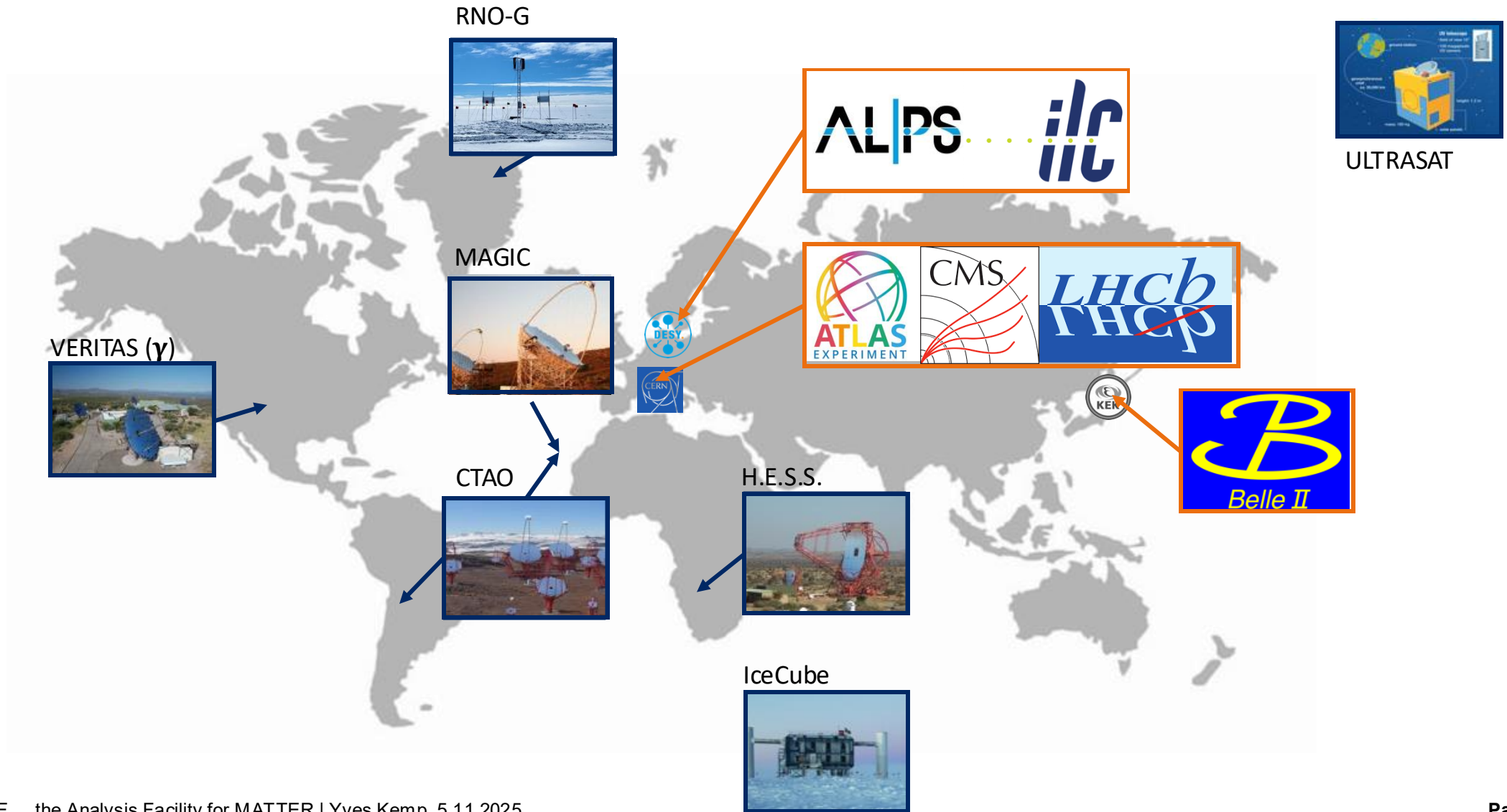


IDAF supports accelerator R & D as well as operation

- Emittance preservation in a plasma-wakefield accelerator, (C. A. Lindstrøm et al., Nature 2024, <https://doi.org/10.1038/s41467-024-50320-1>)
 - Main work carried out on LUMI (TOP500#8)
 - **Preparatory work as well as data analysis on IDAF**
 - Data rates of operational data for PETRA III, EuXFEL, FLASH have increased → **5 PB of metrics store on IDAF**
 - Used e.g. for training LLM (Large Language Model) for autonomous accelerator tuning
- Reinforcement learning-trained optimisers and Bayesian optimisation for online particle accelerator tuning. Jan Kaiser et al. *Sci Rep* **14**, 15733 (2024)



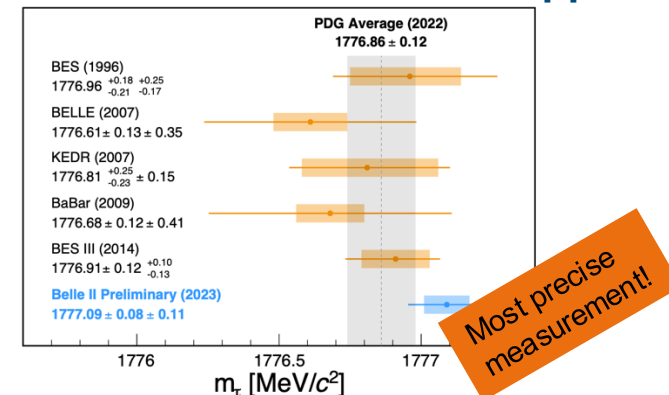
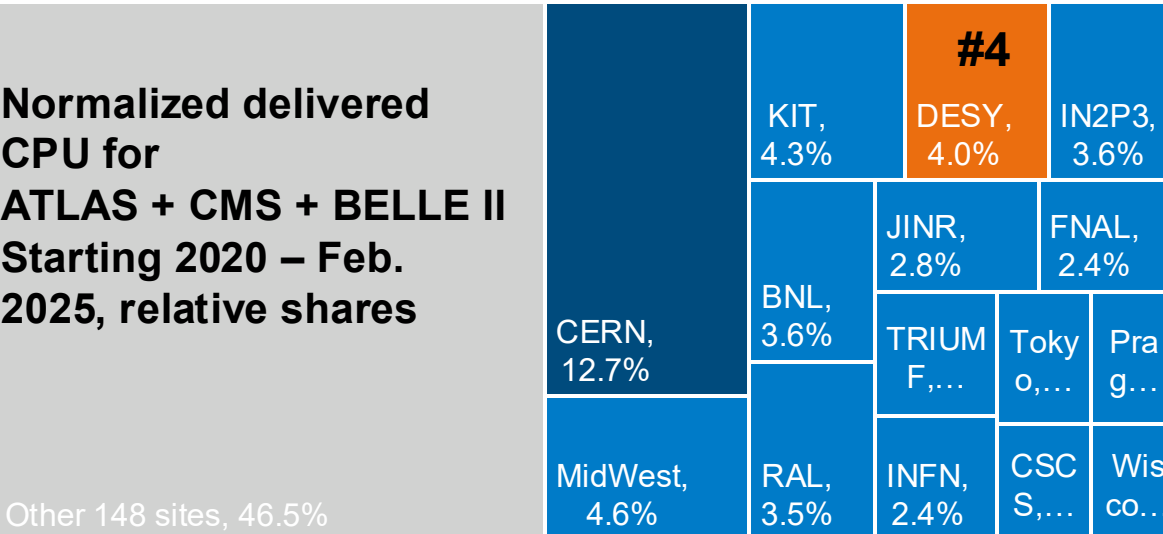
The IDAF supports (Astro-)Particle Physics



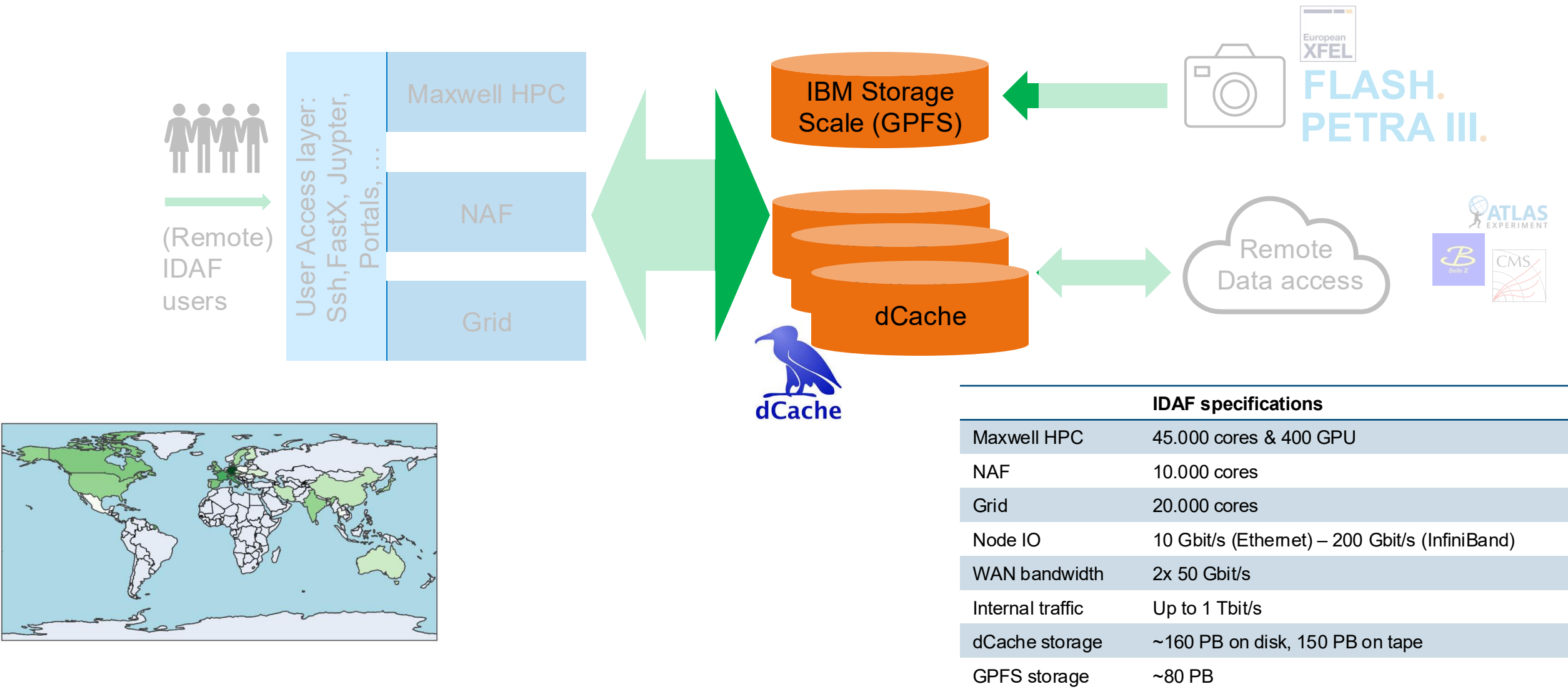
IDAF supports High Energy Particle Physics

- DESY Grid at Hamburg and Zeuthen plays a crucial part as a **large WLCG Tier-2 center**
- **More than WLCG**: Belle 2 RAW data center, one of the offline data center for CTAO
- **Data hosting position** in Germany will be **enhanced** with university Tier-2 fading out

- NAF: National Analysis Facility complements the Grid:
- Federated into experiments data management workflows and software distribution
- Offer **additional services** for accessing the DESY Grid data
- **Interactive Login** and easy, **fast turn-around** compute
- Additional **project space**
- Extensive software, **support and consulting**



The IDAF is data centric compute infrastructure

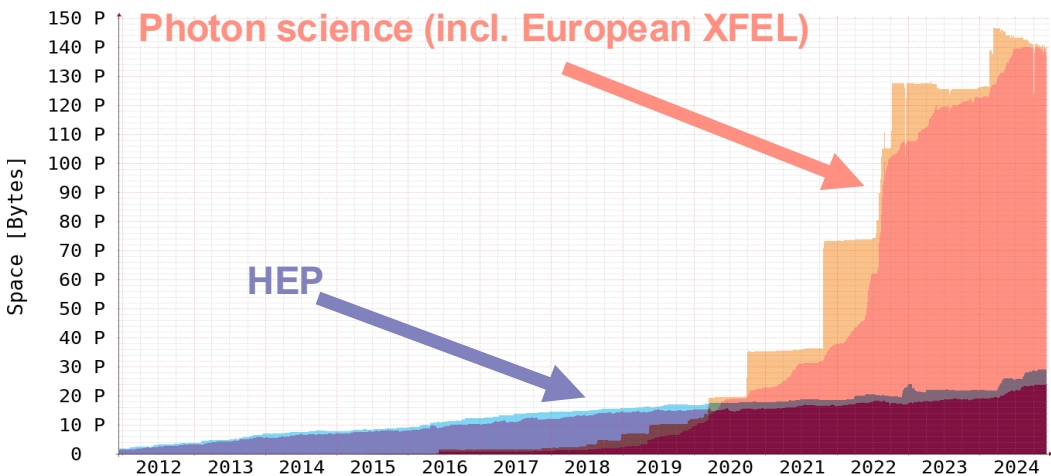


Large, powerful, reliable, high-speed storage systems

dCache:

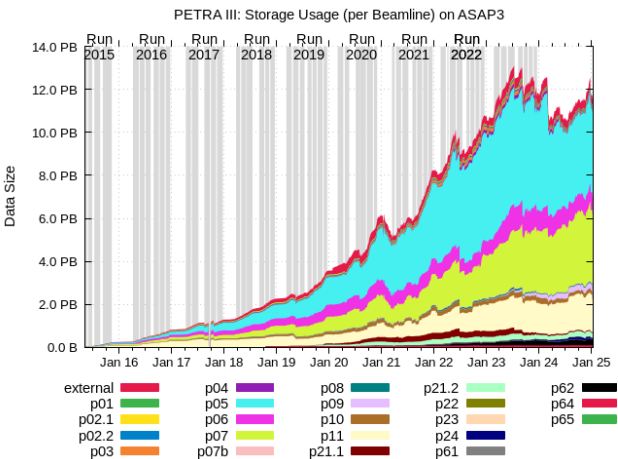
- Worldwide access from different communities via federated methods
- Massive increase for Photon Science basically entirely in PoF-IV period**
 - Data reduction necessary in future

→ **Strong, reliable operation**



IBM GPFS:

- Complements dCache for Single file performance, many small files, meta data heavy
- Since 2014: Strong collaboration with IBM:**
 - Building PETRA III and EuXFEL data taking and analysis on IBM Spectrum Scale / GPFS



- PETRA III : 17 PB
- EuXFEL : 64 PB
- + 5 PB project space

The IDAF is empowered by motivated and skilled people

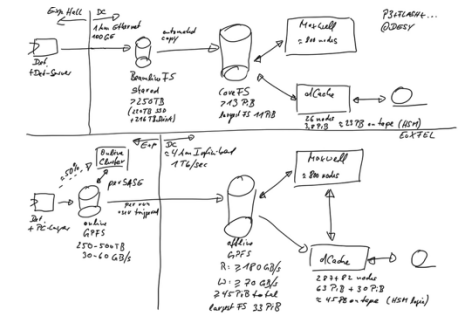
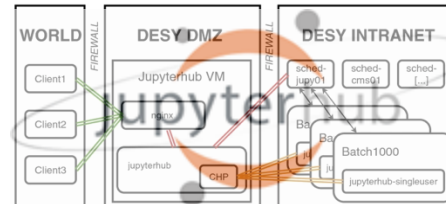
- Support: ~1000 tickets / year
- Training, consulting, strategic planning
- Performing research for the IDAF
- Connecting with (inter-)national initiatives
- Operating the IDAF on a day-to-day basis

```
[rt #1467480] GPU run on tensorflow kernel
```



ALPS poster: Rachel Wolf

Sustainable computing poster:
Eleanor Jones



PETRA III initial DAQ scheme (2014)

Interactive analysis notebooks on
DESY batch resources
Johannes Reppin et al. (2021)
CSBS/Springer



Strategy: build on proven strengths to meet the next generation of data challenges

Strength:

- Proven, reliable user service across diverse communities
- Integrated HTC/HPC/AI infrastructure and support
- Strong integration into communities

Weakness

- Limited personnel capacity for growing user base
- Complex governance and decision processes across communities
- Constrained data-center space and power

Opportunity

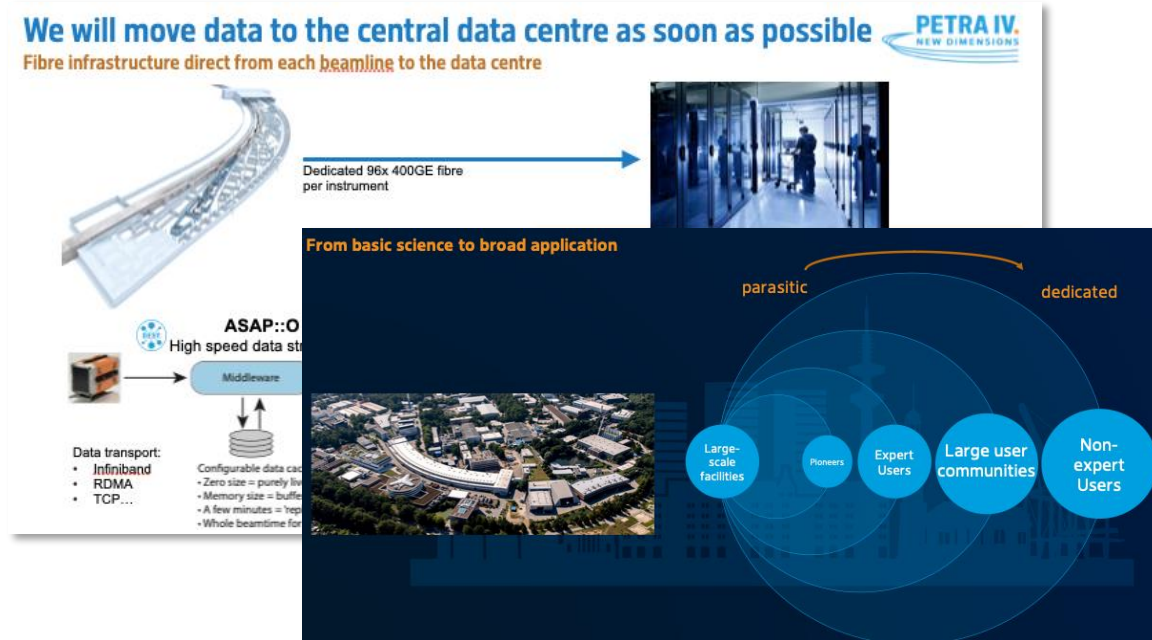
- PETRA IV & HL-LHC: next-generation data processing and FAIR integration
- Expand innovation & partnerships through new communities and industrial partners
- Closer alignment with MT-DMA as well as (inter-)national initiatives

Threats

- HL-LHC & PETRA IV decisions and procedures still in progress
- Rapid tech evolution, procurement risks, and cybersecurity exposure
- Competing priorities among programs and uneven adoption of new standards

Enabling next generation data sources: Petra IV and HL-LHC

PETRA IV: The Analytics Facility

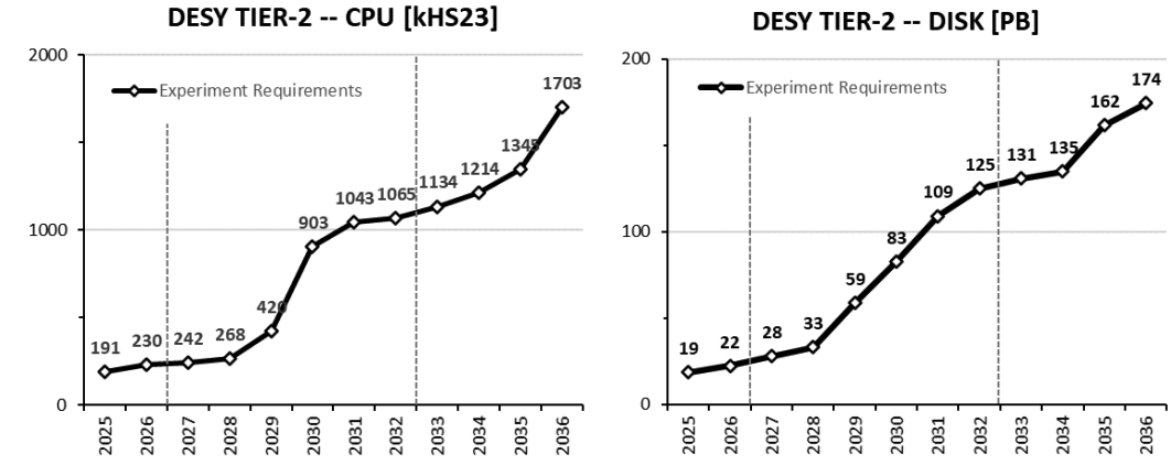


- From Science Facility to Analytics Facility:

Photon Science as a Service

→ Collaboration with MT-DMA and MT-DTS, as well as MML

HL-LHC



- Data increase
 - LHC compute might outstrip PETRA IV
 - IDAF central data hub for all German LHC research
- FIS “TIER-Upgrade”, connection with ErUM for efficient, federated computing.

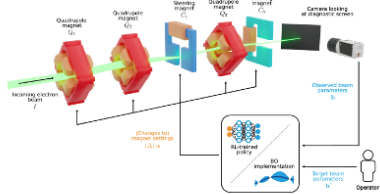
AI & ML: transform analysis, operations, and user experience.

Accelerating science:

- AI-assisted data acquisition, data analysis and data reduction

Data rates of operational data for PETRA III, EuXFEL, FLASH have increased → **5 PB of metrics store on IDAF**

- Used e.g. for training LLM (Large Language Model) for autonomous accelerator tuning
- Reinforcement learning-trained optimisers and Bayesian optimisation for online particle accelerator tuning. Jan Kaiser et al. *Sci Rep* **14**, 15733 (2024)

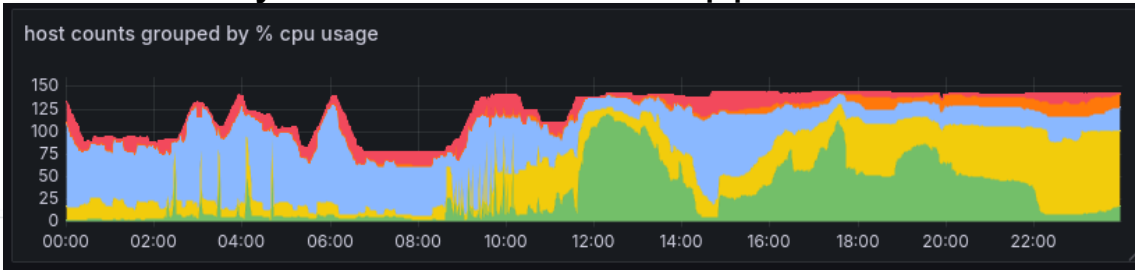


- Integration of AI frameworks within IDAF GPU environments

→ Collaboration with MT-DMA and communities

Enhancing facility operations

- Predictive scheduling, automated maintenance and telemetry. AI assisted user support



TicketNR			
1548944	2025-11-02T13:17:59Z	open	Condor problems
1548635	2025-10-30T13:23:31Z	open	ssh into running jupyter node
1548218	2025-10-29T10:38:04Z	new	NAF/DUST Ressourcen fuer Sustainability Workshop
1547757	2025-10-28T08:21:20Z	open	Sharing dust data with DESY collaborators in other g

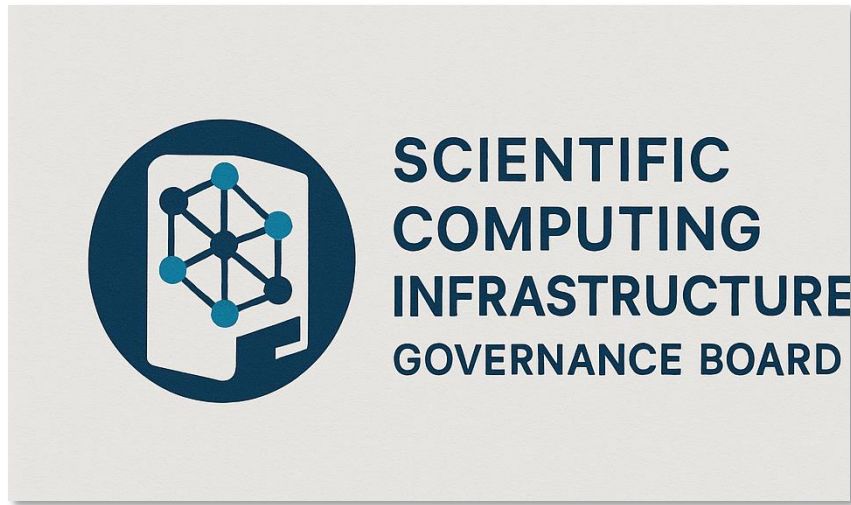
→ Synergy between AI science and IDAF operation

IDAF enables innovation by extending the multi-community approach

- PETRA IV will enlarge focus on innovation case, e.g. extend to industry for data-driven R&D – so will the IDAF
- Provides compliant, auditable pathways without impacting academic workloads
- Operated with DESY ITT and HPC Gateway project to ensure legal and policy alignment
- Strengthens technology transfer and DESY's innovation ecosystem



IDAF integrates communities into governance.

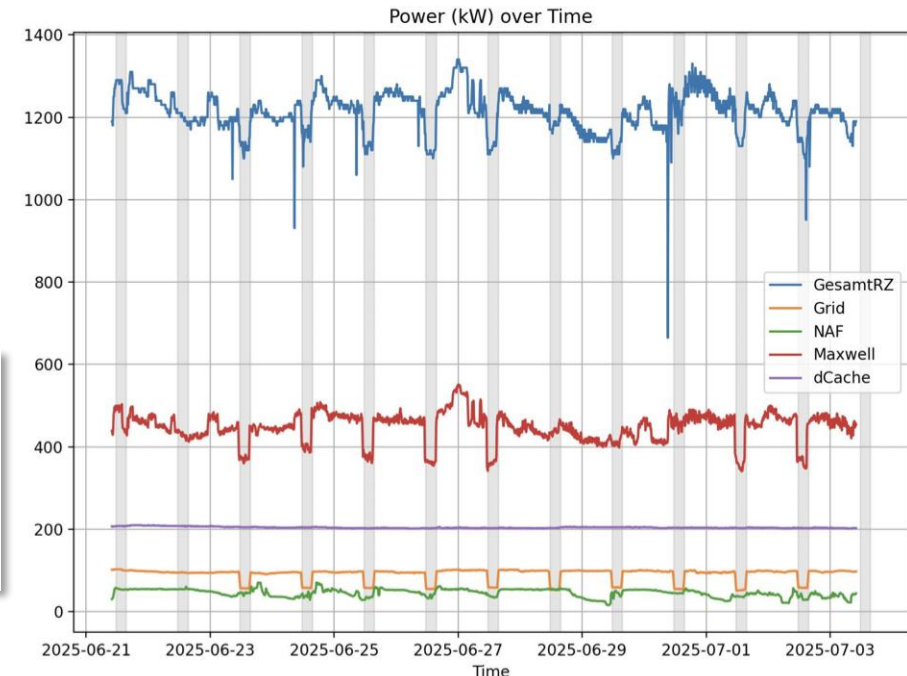
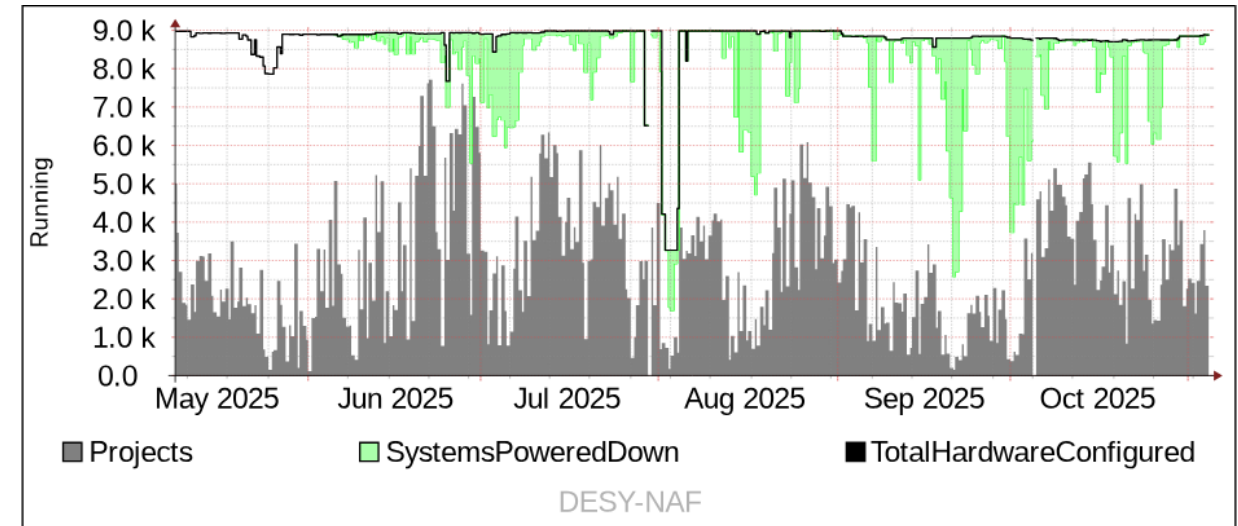


AI generated logo

- Establishes the IDAF Council with technical & scientific representatives from all domains
- Creates dedicated User Committees (Particle, Photon, Machine, Commercial) for coordination
- Ensures shared strategy, service catalogues, and training across communities
- Strengthens collaboration with MT-DMA, NFDI, EOSC, as well as other partner initiatives

Sustainable computing — efficient, resilient, and future-ready.

- Plans for new data-centre capacity with energy-efficient cooling and power design
- Uses telemetry for energy-aware scheduling and predictive maintenance
- Optimizes storage tiers for minimal footprint and maximal reuse
- Optimize hardware life cycle and reduce greenhouse-gas footprint
- Ensures resilience through synergy with Zeuthen and disaster-recovery planning



IDAF: the Analysis Facility for MATTER

Summary: The IDAF:

- a **data-centric**, high-end facility for **high-throughput** and research-oriented **data analysis**
- is well established in **all MATTER programs**
- enables **prime scientific results**

Outlook: The IDAF

- enables **next generation data challenges**
- enables new **AI&ML** workflows and opens to **industry**
- strengthens **user orientation** through new governance
- is well prepared for **future challenges**

To users, the IDAF should remain “black magic”,
that just works for them getting their science done

