HELMHOLTZ

MATTER AND TECHNOLOGIES

RESEARCH FOR GRAND CHALLENGES

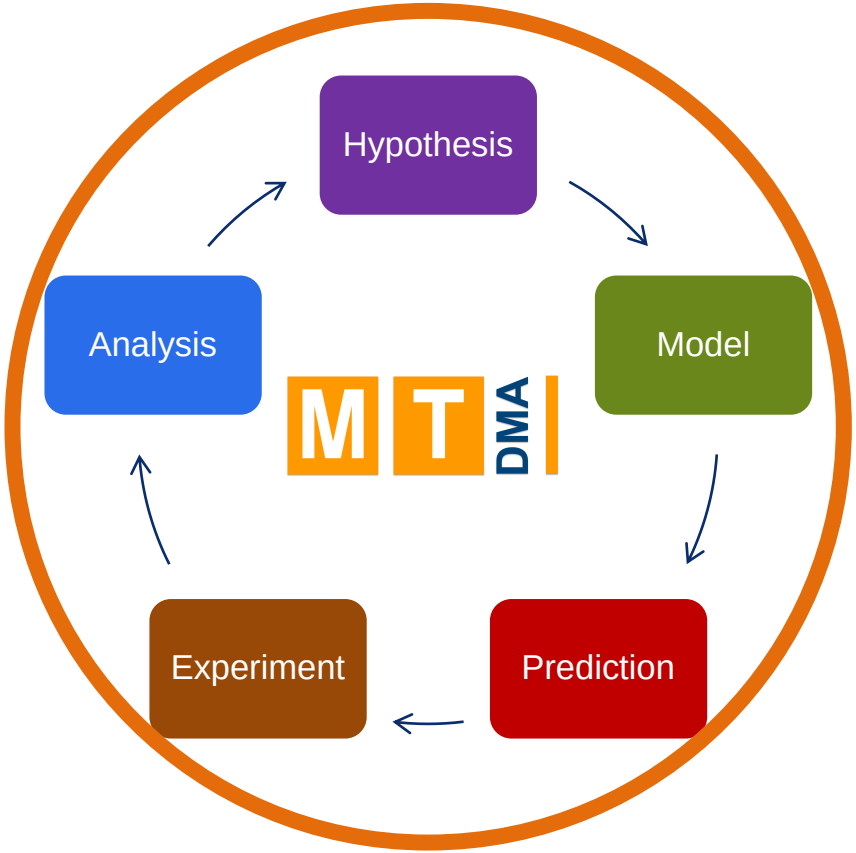
DATA MANAGEMENT & ANALYSIS (DMA)

Michael Bussmann (HZDR)
Volker Guelzow (DESY)

with contributions from
DESY, FZJ, GSI with HI Jena & HI Mainz,
HZB, HZG, HZDR

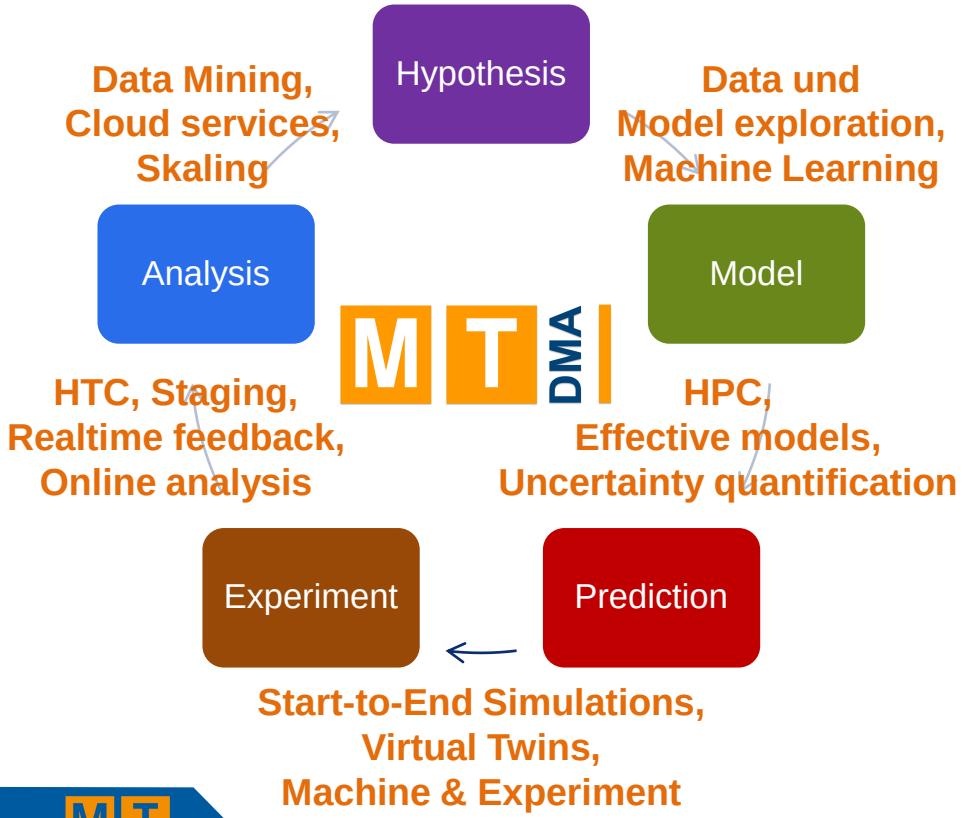
DESY JÜLICH HI Jena Helmholtz Institute Jena HIM Helmholtz Institute Mainz GSI HZB Helmholtz Zentrum Berlin Helmholtz-Zentrum Geesthacht Centre for Materials and Coastal Research HZDR

SCIENTIFIC & DATA-INTENSIVE COMPUTING IN DMA



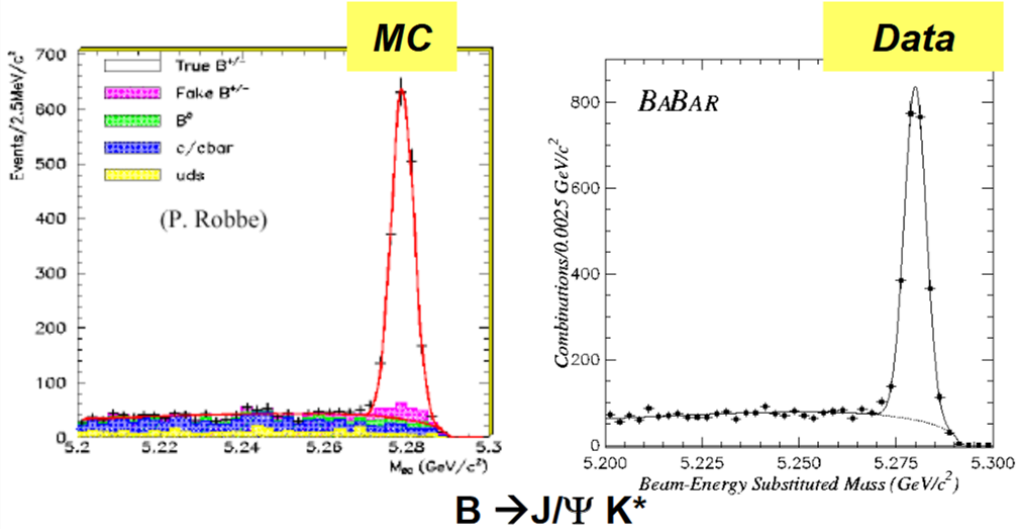
- DMA **integrates technologies, methods and infrastructures** in scientific & data-intensive computing with **research and applications in MATTER,** focusing and joining the various efforst in MATTER.
- DMA includes **all parts of the scientific workflow**, from model building to simulation, from experiment and machine to analysis

SCIENTIFIC & DATA-INTENSIVE COMPUTING IN DMA

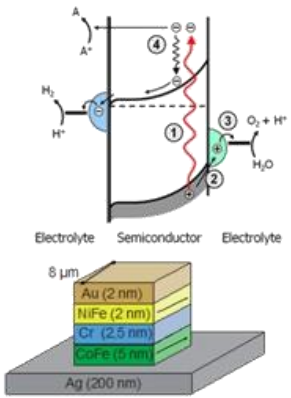


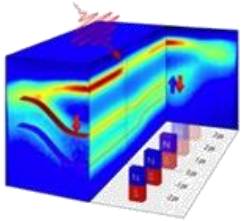
- DMA **integrates technologies, methods and infrastructures** in scientific & data-intensive computing with **research and applications in MATTER,** focusing and joining the various efforst in MATTER.
- DMA includes **all parts of the scientific workflow**, from model building to simulation, from experiment and machine to analysis
- DMA sees the **efficient and complete extraction of knowledge** from data as a **digital process** that involves domain scientists, IT, data scientists and mathematicians, software engineers, machine and detector experts.

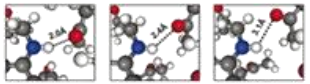
THE HEP LESSON – DIGITAL TWIN EXPERIMENTS

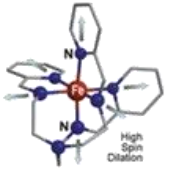


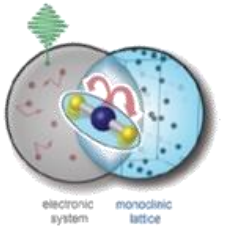
B →J/Ψ K*

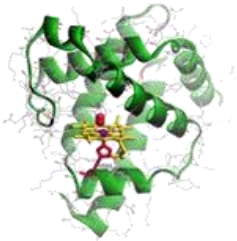


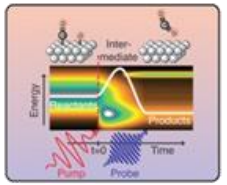










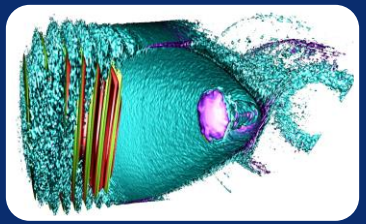


- **Data reduction** essentially relies on **knowledge extraction**
- The optimum case means **comparing** ideal experiments to reality
- Can we build **digital twins** to all experiments in MATTER?

DMA STRUCTURE

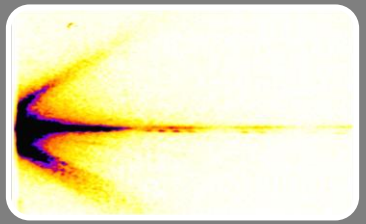
ST1: The Matter Information Fabric

- IT infrastructure (Hard- & Software) for facilities
- Automization of Data Lifecycle Management (LK II)
- Solutions für Communities



ST2: The Digital Scientific Method

- Matter-specific research in Data Analysis & Simulation methods
- e.g. Machine Learning, Simulation, Visual Analytics, Scientific Workflow
- Developing methods für heterogeneous HPC, HTC, I/O for Matter applications

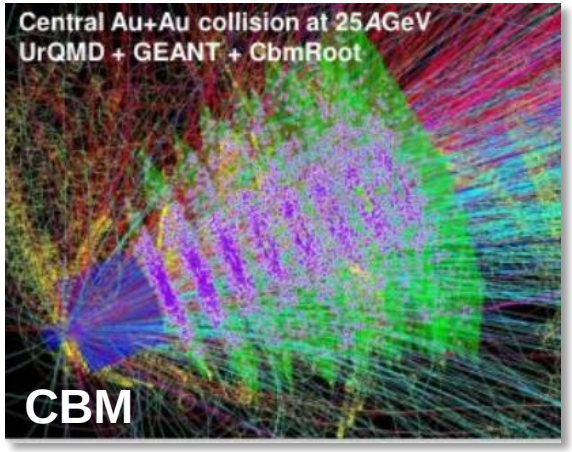


ST3: The Digital Experiment and Machine

- Start-to-End Simulations (Machine/Interaction/Detectors)
- Fast feedback & machine control („Human in the Loop“)
- Quantifying data quality, meta data acquisition & analysis

ST1 MATTER INFORMATION FABRIC

- **Infrastructure (Soft- & Hardware)** for Data & Meta Data Lifecycle Management at LK-II facilities
- Solutions & technologies for large-scale **data ingest, transport & analysis**
- Integration of **machine control**
- Creation of **community platforms**
- **Cross-community** solutions
- **Providing the fabric** for high-throughput, high performance data handling

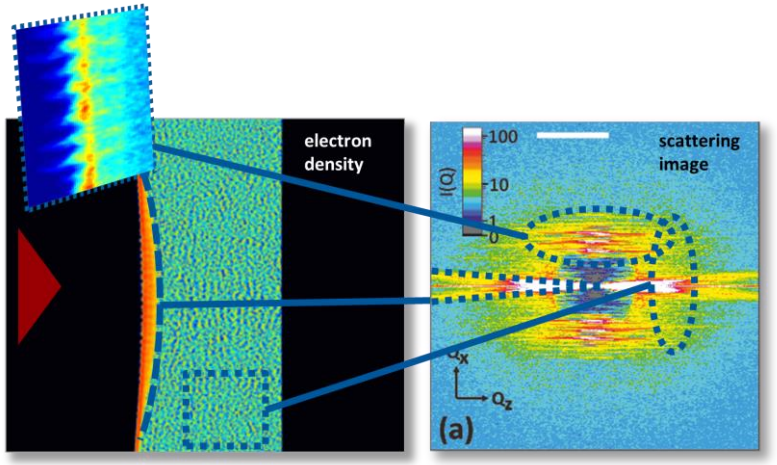


ST1 CHALLENGES

- **Integration** of solutions for high-throughput, high-volume, high-complexity demands at large-scale facilities (DAQ, Control, Ingest, Transport, Storage, …)
- **Cross-center software** development, interoperability & sharing
- **Cross-community software** development, interoperabilty & sharing
- **Cross-center data object** definiton & sharing following FAIR principles
- **Reusing** parts of large LK-II facility platforms for diverse & small-scale communities

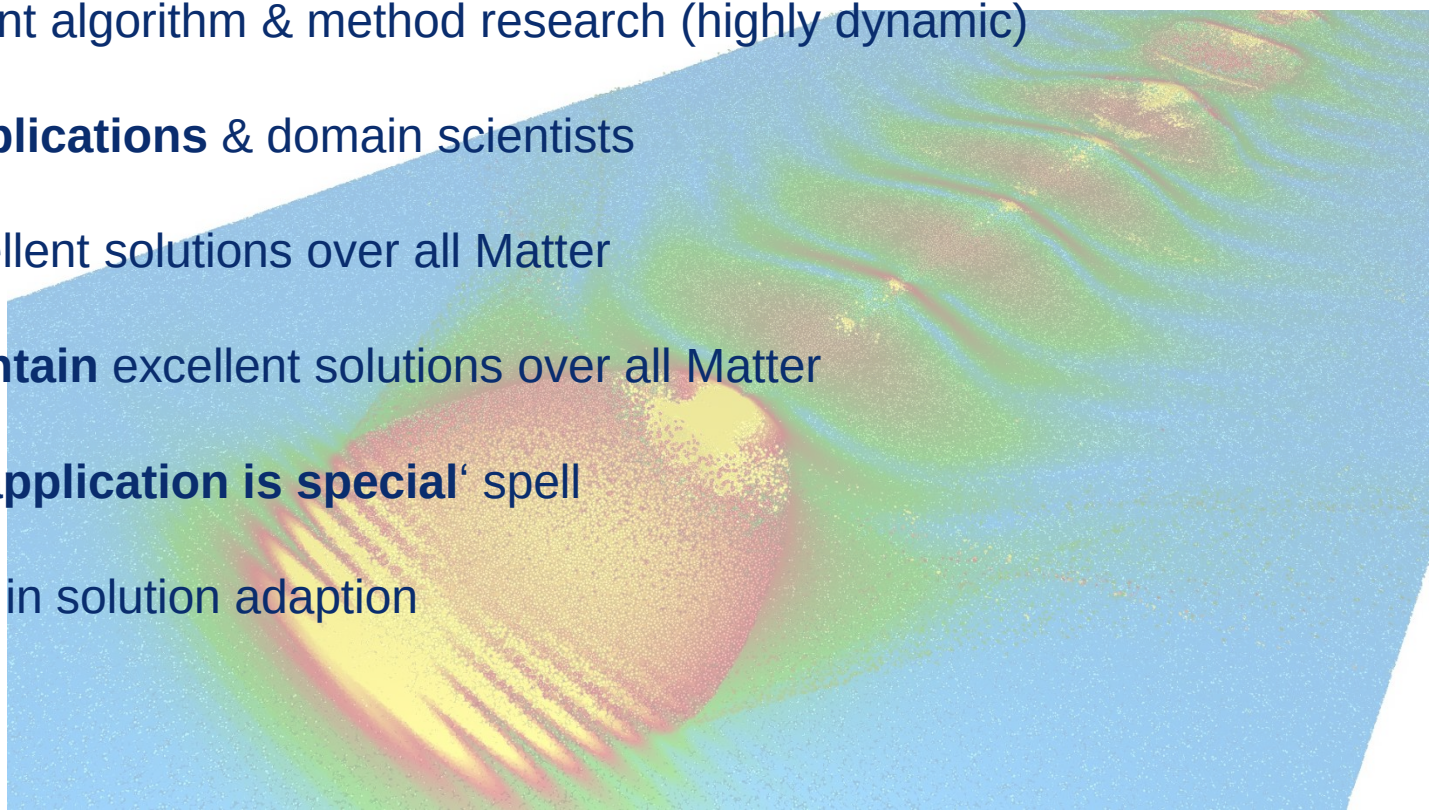
ST2 THE DIGITAL SCIENTIFIC METHOD

- **Excellent methods for excellent science**
- **Method & technology research and development** for research within Matter
- Method & technology research is **application-driven**
- **Sharing methods & solutions between application domains**
  (e.g. particle physics & photon science)
- **Fusion of experimental & model data** via **scalable** methods (HPC, HTC, ML,…)



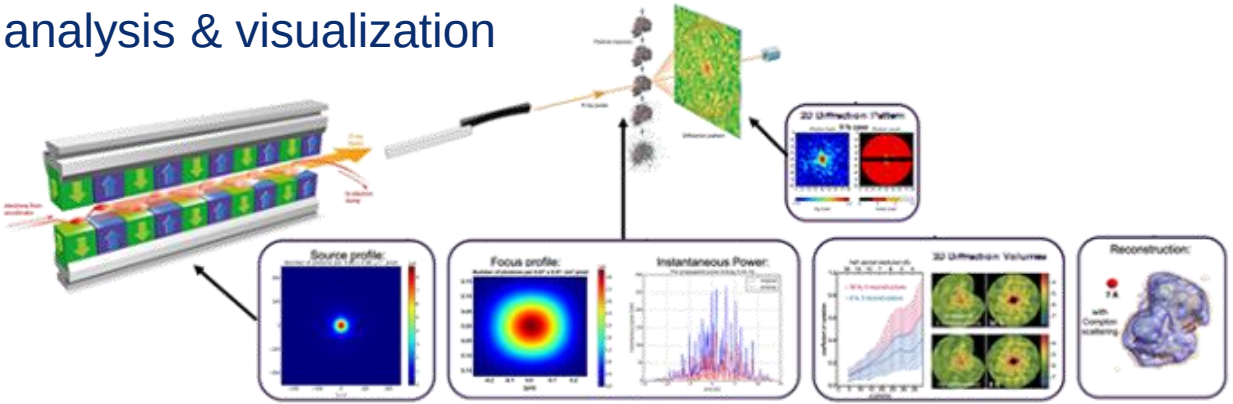
ST2 CHALLENGES

- **Identify** excellent algorithm & method research (highly dynamic)
- **Connect to applications** & domain scientists
- **Distribute** excellent solutions over all Matter
- **Provide & maintain** excellent solutions over all Matter
- Break the ‚**my application is special**‘ spell
- **Educate users** in solution adaption



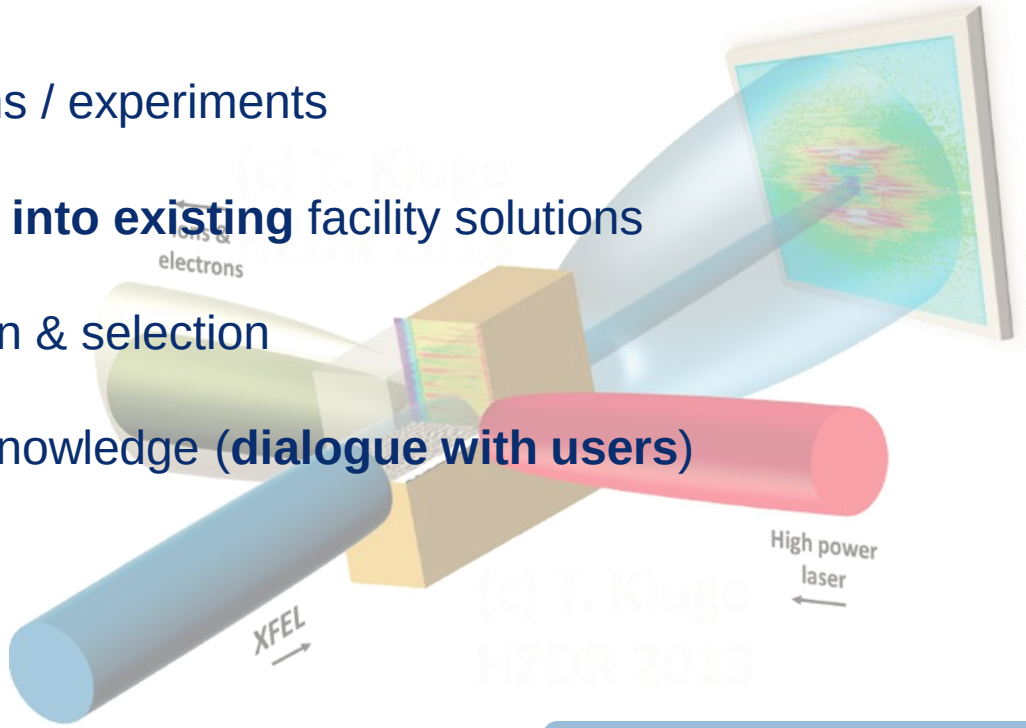
ST3 THE DIGITAL EXPERIMENT & MACHINE

- **Start-to-end simulations**, synthetic diagnsotics & digital twins
- **Fast feedback**, experimental **&** machine **control**
- Data **quality** assessment, **meta data** acquisition, **open data formats**
- **Experiment-specific / community-specific high-throughput / -volume** workflows for (in-situ) analysis & visualization
- Learn from HEP

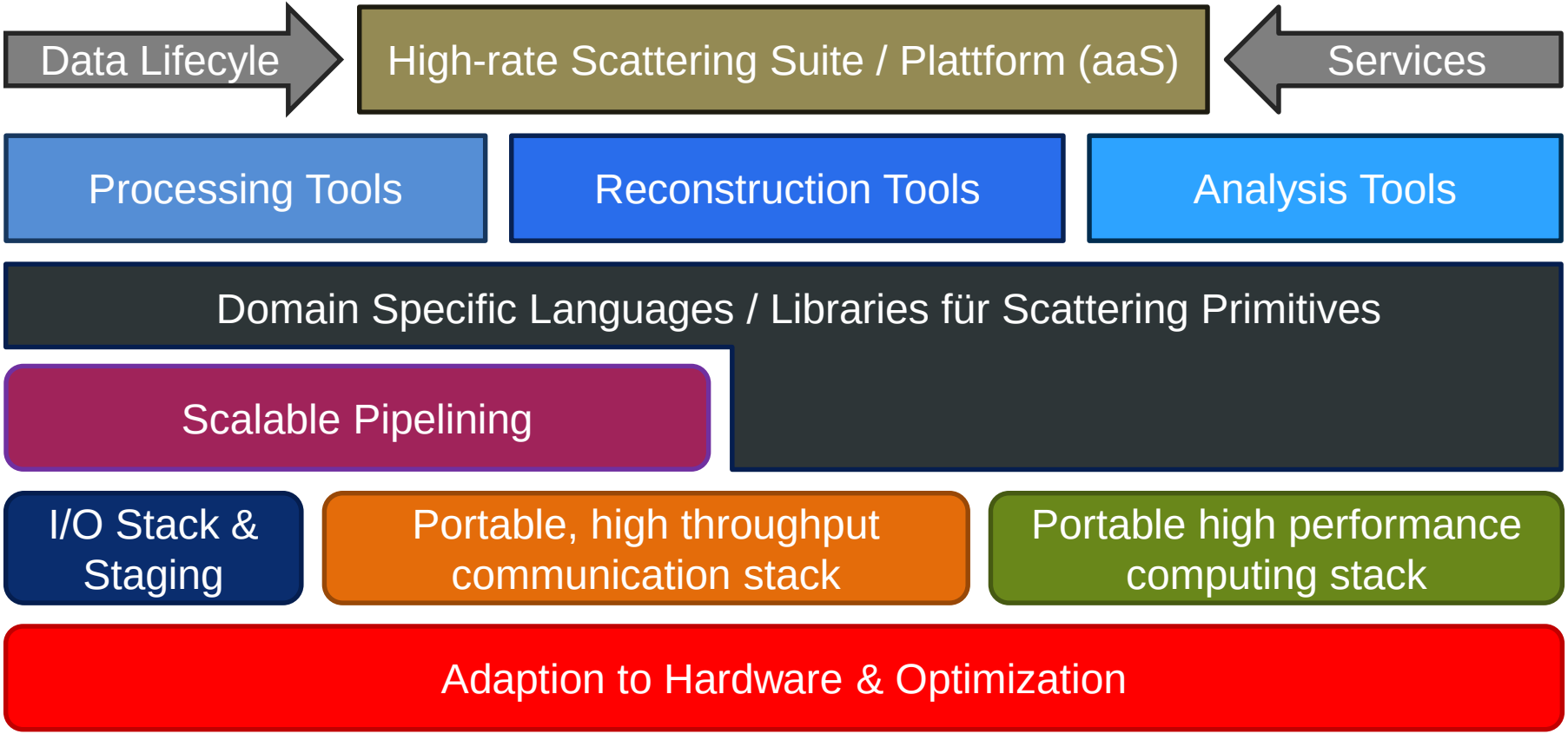


ST3 CHALLENGES

- **Providing open platform solutions** (DAQ, control, analysis) reusable beyond a single facility
- **Adaption to individual** endstations / experiments
- **Integrating indiviudal workflows into existing** facility solutions
- **Data object** identification, definition & selection
- **Data reduction** requires domain knowledge (**dialogue with users**)



MODULAR, REUSABLE, SHAREABLE SOLUTIONS



MATTER MEETS INFORMATION

Excellent driving applications
(3D+T X-ray Imaging, Plasma-based Accelerators, Matter & Materials, …)

Advanced methods for modelling & data analysis
(HPC Start-to-end Simulations, data-driven modeling, data fusion, ML+Simulation, …)

Scalable technologies for high performance & throughput demands
(Heterogeneous computing, Transport & I/O, Workflows, Meta Data, Cloud, …)