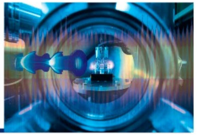
HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES

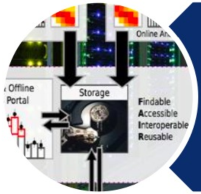
MT-DMA-ST2 Perspective

Mohammad Al-Turany
Guido Juckeland

DMA — Set up for the digital future

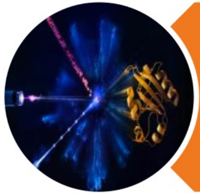


DMA creates new synergies between centers, facilities, communities & leverages them



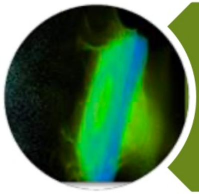
ST1 — The Matter Information Fabric

- Exascale Data Management
- F.A.I.R. Data & Meta Data
- Long-term preservation



ST2 — The Digital Scientific Method

- Artificial Intelligence
- Exascale Computing
- Near real time analysis
- Quantum Computing

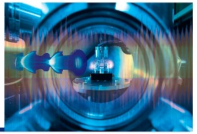


ST3 — The Digital Experiment and Machine

- Exascale Simulations
- In situ Data Analytics
- Near real time feedback
- Machine optimization

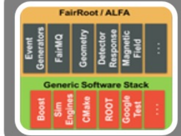
regular exchange & common projects

DMA ST 2 — The Digital Scientific Method



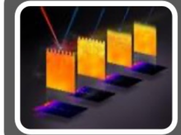
2023

- DMA repository of interconnectable, modular software in full operation



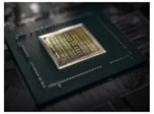
2025

- Toolbox for near-realtime data analysis at extreme scales available

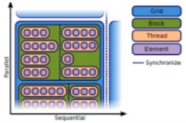


2027

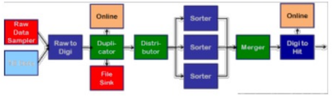
- Surrogate models of multi-source, multi-modal experiments



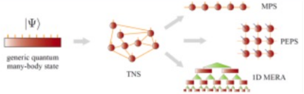
Scalable GPU, FPGA, TPU algorithms



Perfromance-portable programming



Scalable data distribution



Quantum computing for QED

Near real time A.I. analysis of complex systems

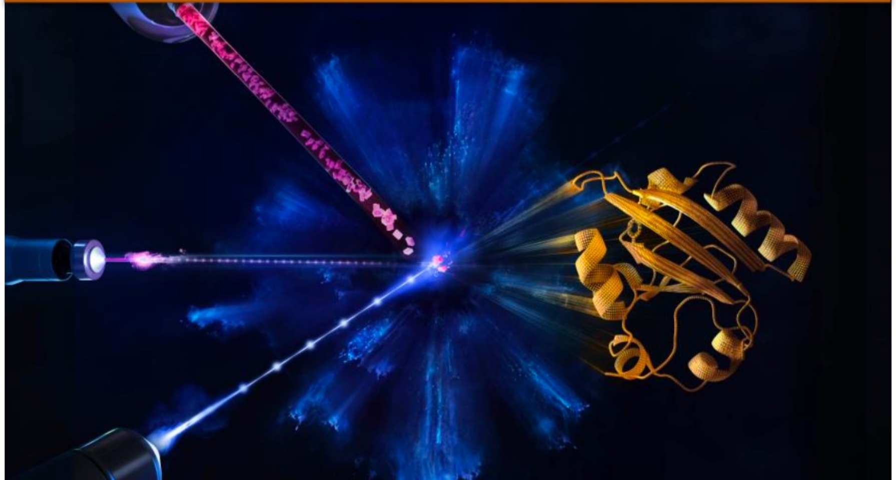
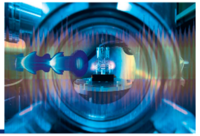


Image: Time-Resolved Serial Femtosecond Crystallography

- Artificial Intelligence
- Exascale Computing
- Near real time analysis
- Quantum Computing

Staying ahead of the data deluge



Milestone		Subtopic	Year
DMA-4	Workshop defining and strengthening synergies in data lifecycle management among the participating facilities and communities	ST1	2022
DMA-7	Provide a directory of interconnectable software packages including examples to cover the whole simulation and experiment life cycle	ST2	2023
DMA-2	S4M portal goes online	ST 1-3	2024
DMA-8	Integration of near-real-time/online data analysis solutions for extreme scale data into the software toolbox of DMA	ST2	2025
DMA-12	Successful demonstrators of digital twins providing virtual data sets mimicking real time, real data operation	ST3	2027
DMA-3	All DMA-provided solutions are available online via S4M	ST 1-3	2027

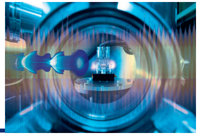
New links created

Key capabilities available

Broad impact on all scales

Continuous delivery of solutions following facilities' schedules

Main Challenges from ST2 → ST1



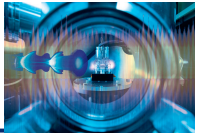
How to make software FAIR?

- working on "F" first
- increasing recognition of software

How to integrate simulation data?

- Data formats/standards
- Data provenance for very large datasets

DMA ST 2 — Towards Milestone DMA-7 (2023)



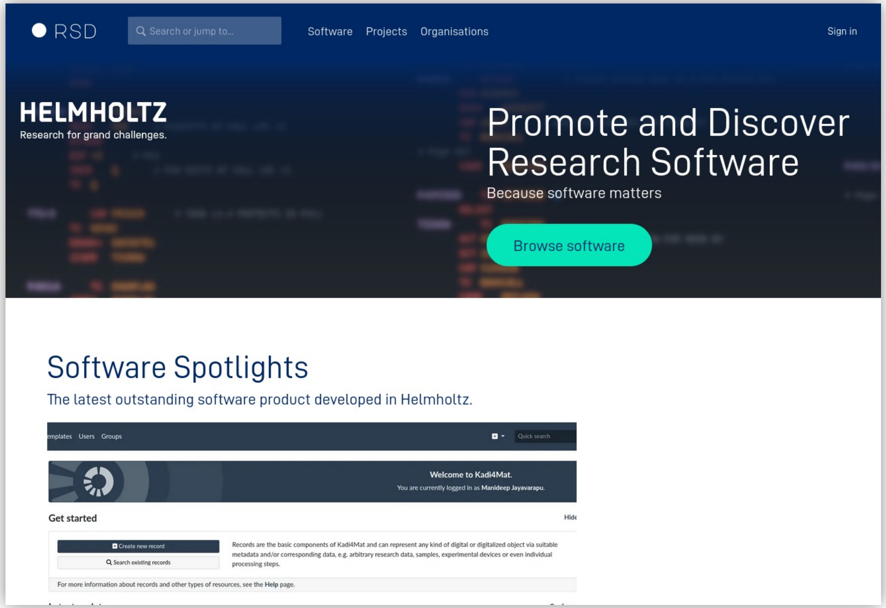
- **Provision of a directory of interconnectable software packages including examples to cover the whole simulation and experiment life cycle**
  - ESCAPE-OSSR: provides a very good curation workflow, actual directory is a repository (https://zenodo.org/communities/escape2020/search?q=&type=software)
  - Helmholtz Research Software Directory: highly flexible software directory, but (currently) no curation

    (https://helmholtz.software)
  - Idea: Combine workflows from both -> Use automated tools

ESCAPE ESCAPE OSSR Open-source Scientific Software and Service Repository

Research Software Directory

Aim and benefits



https://helmholtz.software

- Online service to collect and present software in an **academic context**
- For **Research Software Engineers**
  - Show impact their software has in research
  - Show relations to organisations, research projects and other software
  - Guide visitors to codebase
- For **Researchers**
  - Discover software they need in their research field
  - Get help for citing code they use
- For **Organisations**
  - Keep track of software
  - Metrics and evaluation

Data sources

Authentication

HELMHOLTZ AAI

Code information

GitLab

GitHub

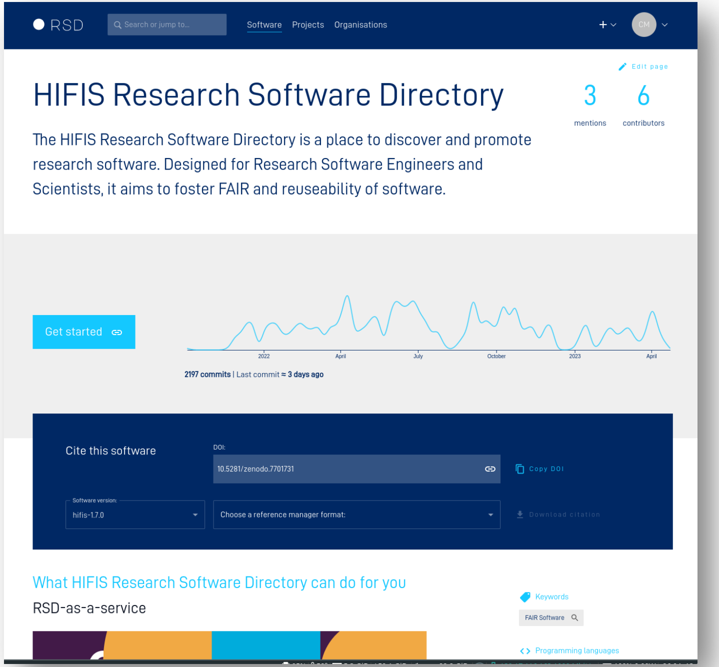
python Package Index

ANACONDA

docker hub

Maven

npm



Organisations

ROR

Contributors

ORCID

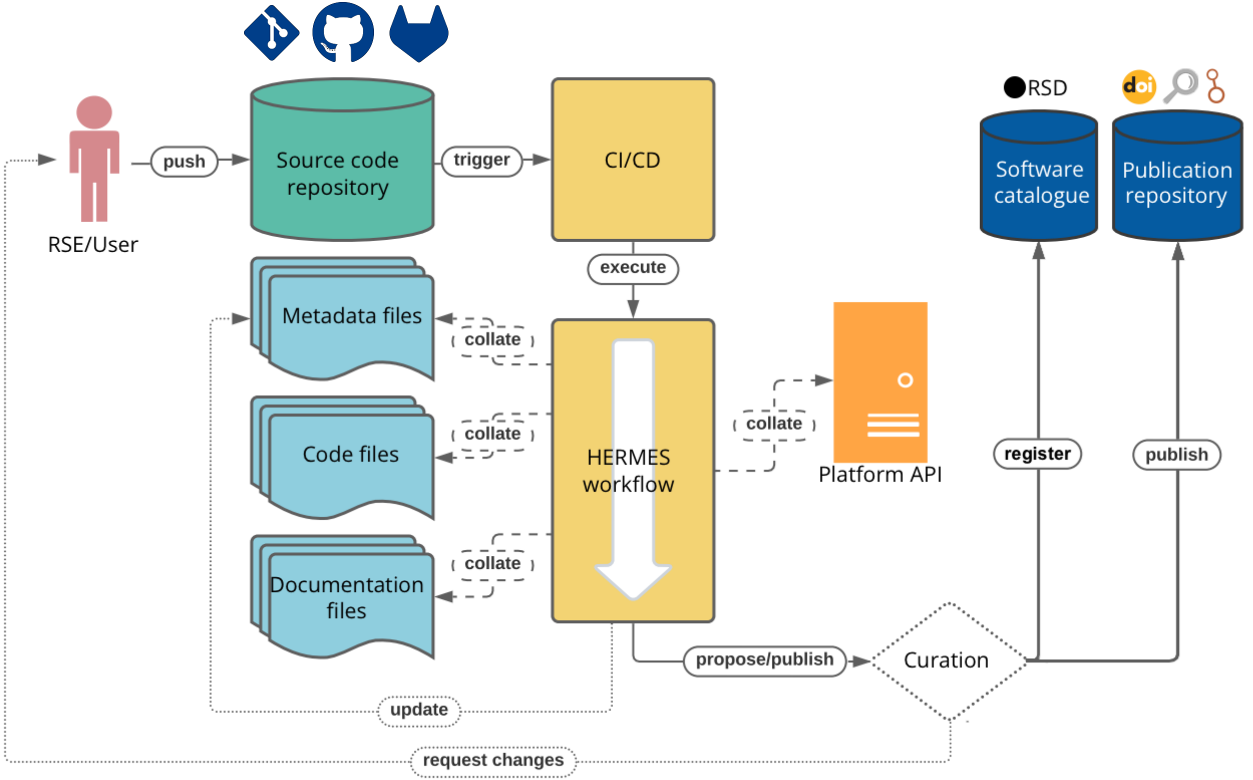
References

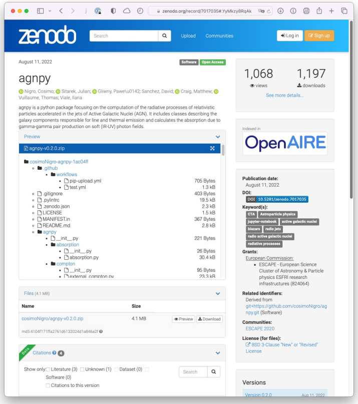
Crossref

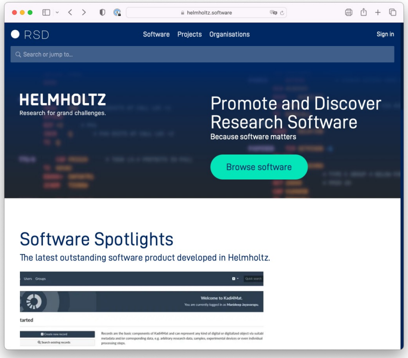
DataCite
FIND, ACCESS, AND REUSE DATA

The HERMES Workflow

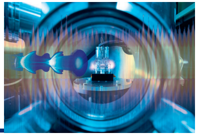
…with Research Software Directory (RSD) Extension

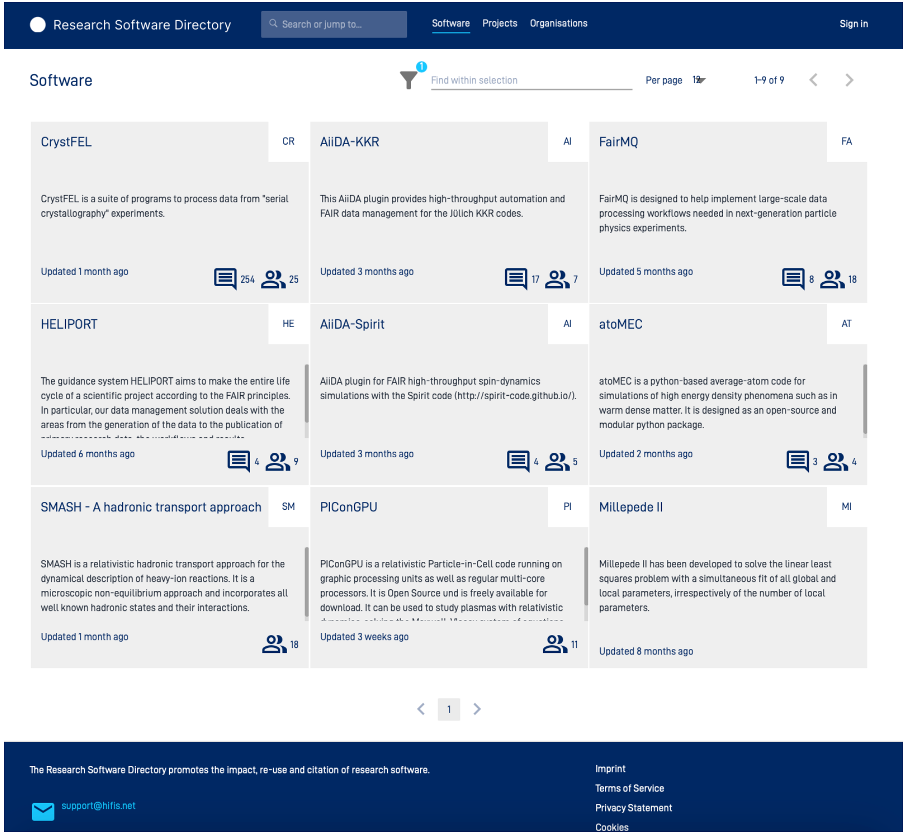






RSD + Matter – current state





- Foundation for Milestone DMA.07 done
- Content not yet rich
- → to be part of S4M portal (DMA.02+.03)

- We need to collect the groups to approach to help onboard their software packages (maybe also now already done via Helmholtz Software Award)
- Regular meetings for onboarding packages

Towards a quality indicator for research software -Status April 2023-

Subgroup Software Quality Indicator

Workpackage 1: Define quality dimensions and attributes

Scientific software of high standard should be reliable & sustainable

=> Quality dimensions to determine „reliable“ and „sustainable“

Quality Dimensions

reliable & sustainable scientific software has to be

- Findable
- Accessible
- Interoperable
- Reusable
- Scientifically well-grounded
- Technologically well-grounded

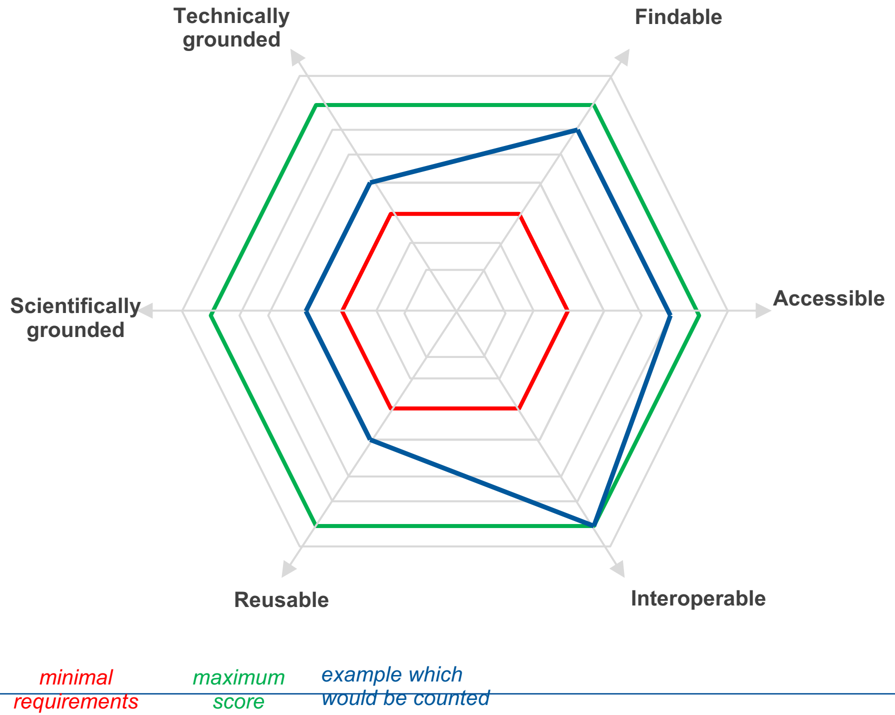
=> FAIR+ST

Scientifically well-grounded means that scientific software is based on scientific knowledge and practice

Technologically well-grounded means that scientific software is based on software engineering knowledge and practice

Workpackage 2: Define maturity levels for quality attributes

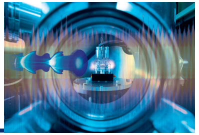
Methode: Multivariate starplot and process-oriented maturity levels, COBIT Maturity Model
(COBIT is an international recognized framework for IT Governance, it is directed to processes)



- Quality Dimensions (FAIR+ST)
- Attributes describing each dimension*
- Maturity levels for each attribute**
- Maturity levels for each dimension** derived form attribute maturity levels

* Paper FAIR4RS DOI: https://doi.org/10.15497/RDA00068
** COBIT Maturity Model

Dealing with data from/with software



Data formats/standards

- Nexus / OpenPMD in Laser Physics
- joint metadata catalogues between simulation and experiments

Data provenance

- How to include simulation workflows (reproducibility)? → ST3
- How to handle multi-TB data sets from photon science?
  - What can we learn/reuse from HEP?
  - What can we learn from other communities?