Three years of DMA-ST2 "The Digital Scientific Method"

Achievements, Lessons learned, Future plans – A personal perspective

Department of Information Services and Computing · Computational Science Department · Dr. Guido Juckeland · g.juckeland@hzdr.de · www.hzdr.de

Mission Statement

DMA-ST2: The Digital Scientific Method

DMA-ST2 develops and maintains a toolbox of

- highly scalable,
- interoperable,
- and community driven

open-source software packages to enable

- leading edge simulations and
- data analysis methods.

Common Interests

Different Domains/Applications, similar Methods

Next generation computing for simulation & analysis

- fast changing compute platforms
- highly scalable software required

Complex data analysis & data fusion

- online reconstruction/data reduction
- automated and/or in-situ

Applications of Machine Learning

- faster convergence of iterative methods
- surrogate models

Knowledge Extraction & Data Reduction

- keeping the "right" <1%

High Throughput Transport

- solutions for short/medium/long range transport
- provide good data formats

DRESDEN
concept

HZDR

Planned steps

ST2 Milestones

Make methods/software visible

- DMA-7: Provision of a directory of interconnectable software packages including examples to cover the whole simulation and experiment life cycle

Make methods/software accessible/exchangeable

- DMA-8: Integration of near-real-time/online data analysis solutions for extreme scale data into the software toolbox of DMA

Connecting to the other STs

- DMA-9: Integration of surrogate models into simulation of multisource, multimodal experiment setups

First realizations: A multitude of software (incomplete list)

Collaboration via knowledge exchange not so much code exchange

CP2K

CP2K is a quantum chemistry and solid state physics software package running on graphics...

Electronic Structure　GPU

High Performance Computing　+ 5

Fortran　C　Python
+ 13　　8　　2143

CORSIKA

A Monte Carlo framework for the simulation of particle showers initiated in the Earth's atmospher...

Astroparticle Physics　Matter

Monte Carlo Simulation

2　　1574

CrystFEL

CrystFEL is a suite of programs to process data from "serial crystallography" experiments.

Crystallography　Matter

Structural Biology　+ 2

C　Julia　Python　+ 2　　25　　341

FairMQ

FairMQ is designed to help implement large-scale data processing workflows needed in...

Framework　Matter

Workflow Library

C++　CMake　C　+ 2　　18　　8

Kadi4Mat

Kadi4Mat is a generic and open source virtual research environment.

Aeronautics, Space And Transport

Earth & Environment

Electronic Lab Notebook　+ 9

Python　Vue　HTML　+ 2　　4　　7

AiiDA-Spirit

AiiDA plugin for FAIR high-throughput spin-dynamics simulations with the Spirit code...

Data Managment　Data Visualization

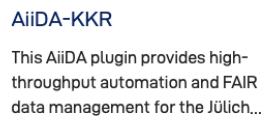FAIR Data　+ 5

JavaScript　Python
HTML　+ 1　　5　　4

Millepede II

Millepede II has been developed to solve the linear least squares problem with a simultaneous fit ...

Alignment

Large Linear Equation System　Matter

Fortran Free Form　C
Python　+ 2　　0　　0

AiiDA-KKR

This AiiDA plugin provides high-throughput automation and FAIR data management for the Jülich...

Data Visualization　FAIR Data

Information　+ 4

Python　Shell
OpenEdge ABL　　7　　17

Propulate

An asynchronous evolutionary optimization algorithm and software package for global...

Aeronautics, Space And Transport

Earth & Environment　Energy　+ 9

Python　　6　　11

HELIPORT

The guidance system HELIPORT aims to make the entire life cycle of a scientific project according t...

Data Managment　Earth & Environment

Energy　+ 10

Python　HTML
JavaScript　+ 2　　9　　9

atoMEC

atoMEC is a python-based average-atom code for simulations of high energy densit...

Matter
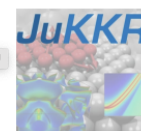
Python　Shell　　4　　3

PIConGPU

PIConGPU is a relativistic Particle-in-Cell code running on graphic processing units as well as regul...

CPU　GPU

Heterogeneous Computing　+ 3

C++　Python　Shell　+ 5　　11　　3

JuKKR

The Korringa-Kohn-Rostoker (KKR) Greens function method is a highly accurate all-electron...

CPU　Density Functional Theory

High Performance Computing　+ 5

Fortran Free Form
Fortran　TeX　+ 2　　29　　0

SMASH - A hadronic...

SMASH is a relativistic hadronic transport approach for the dynamical description of heavy-i...

CPU　Matter　Open Source　+ 1

C++　CMake　Shell　+ 3　　18　　0

Contributing to and benefitting from community projects
Connecting DMA with EOSC and HIFIS

**DMA Software directory as part of larger efforts**

- ESCAPE-OSSR: provides a very good curation workflow, actual directory is a repository (https://zenodo.org/communities/escape2020/search?q=&type=software)

- HIFIS Research Software Directory: highly flexible software directory, but (currently) no curation (https://helmholtz.software)

Current work:

- Combine workflows from both to onboard software into DMA "community" of RSD

- Then export this part of RSD to https://helmholtz-dma.de website

Helmholtz Research Software Directory

Aim and benefits
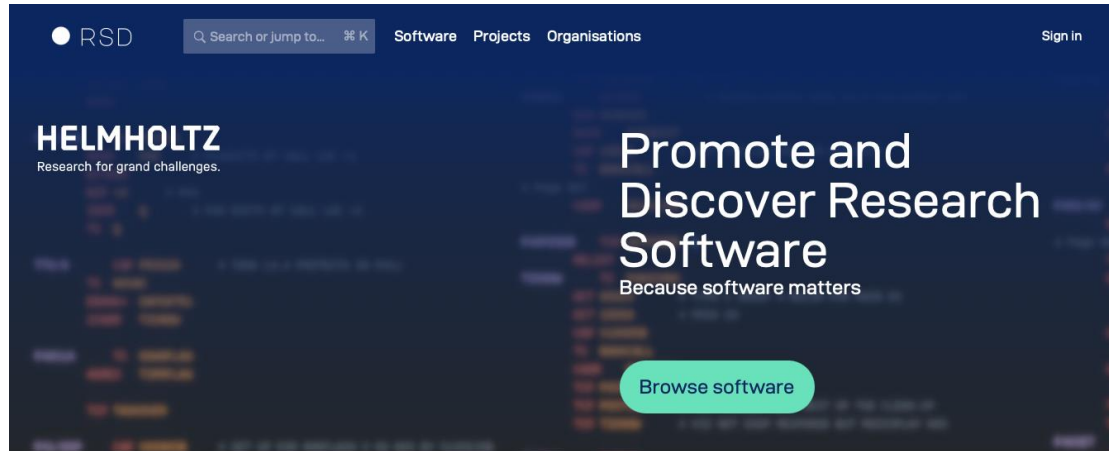
https://helmholtz.software

- 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

RSD Data sources

Organisations

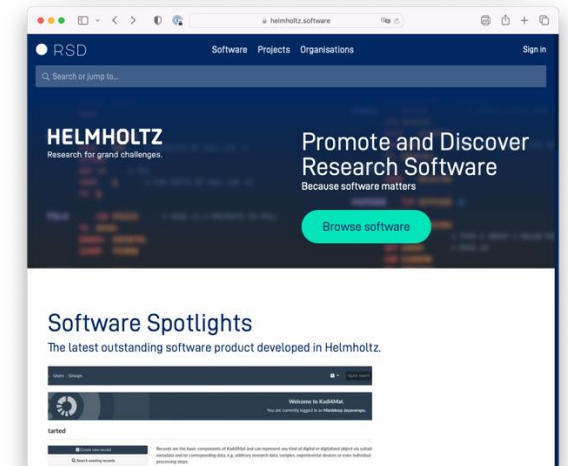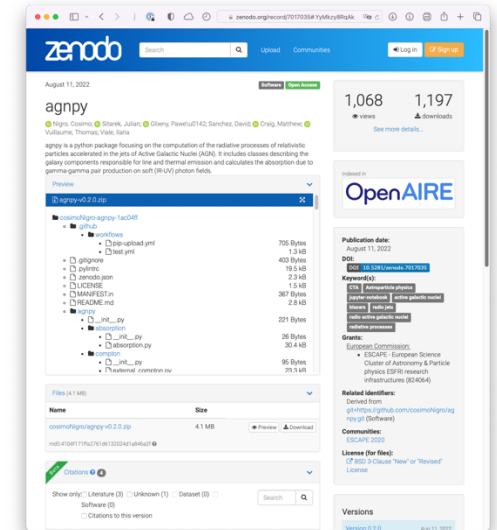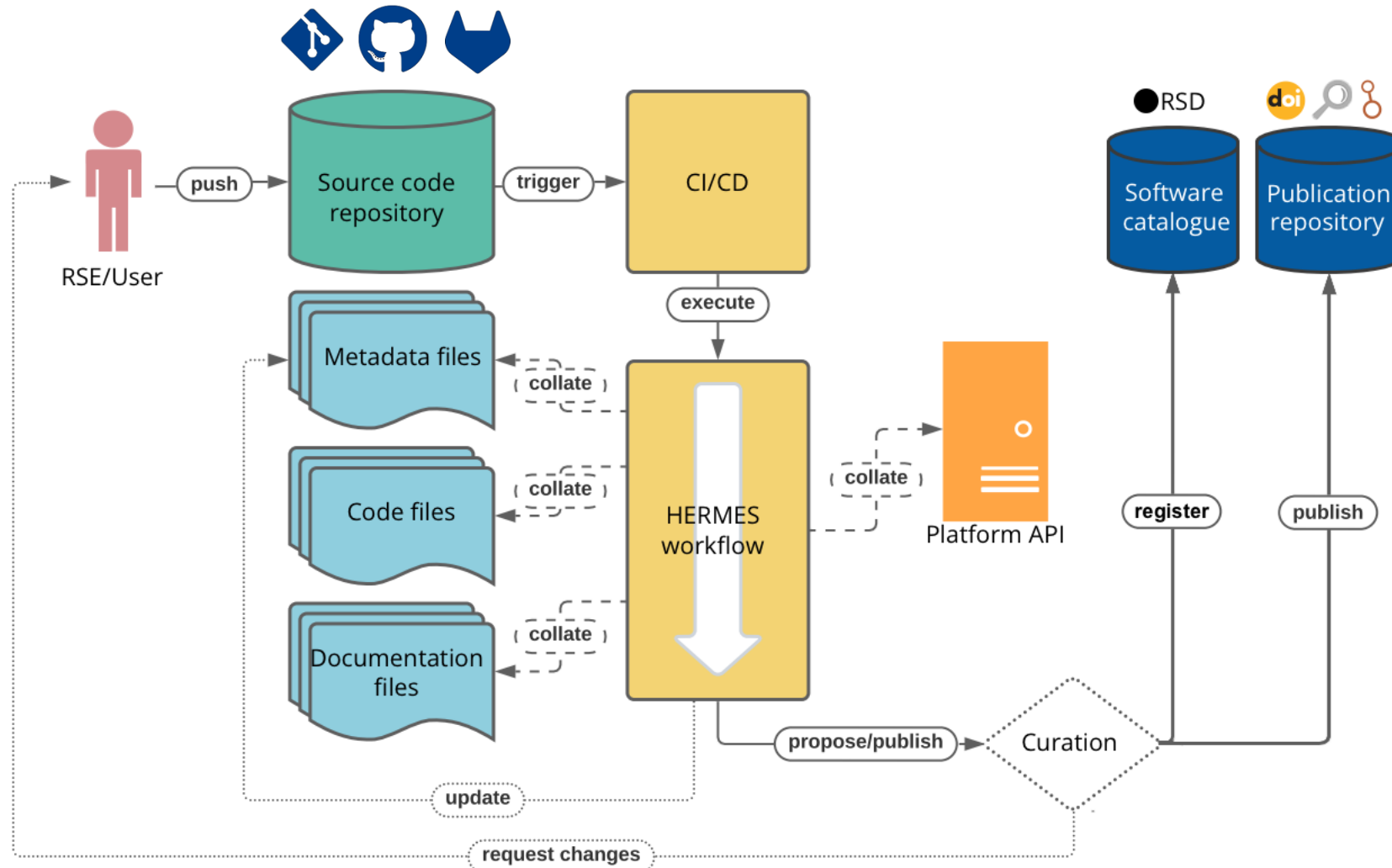Authentication

Code information
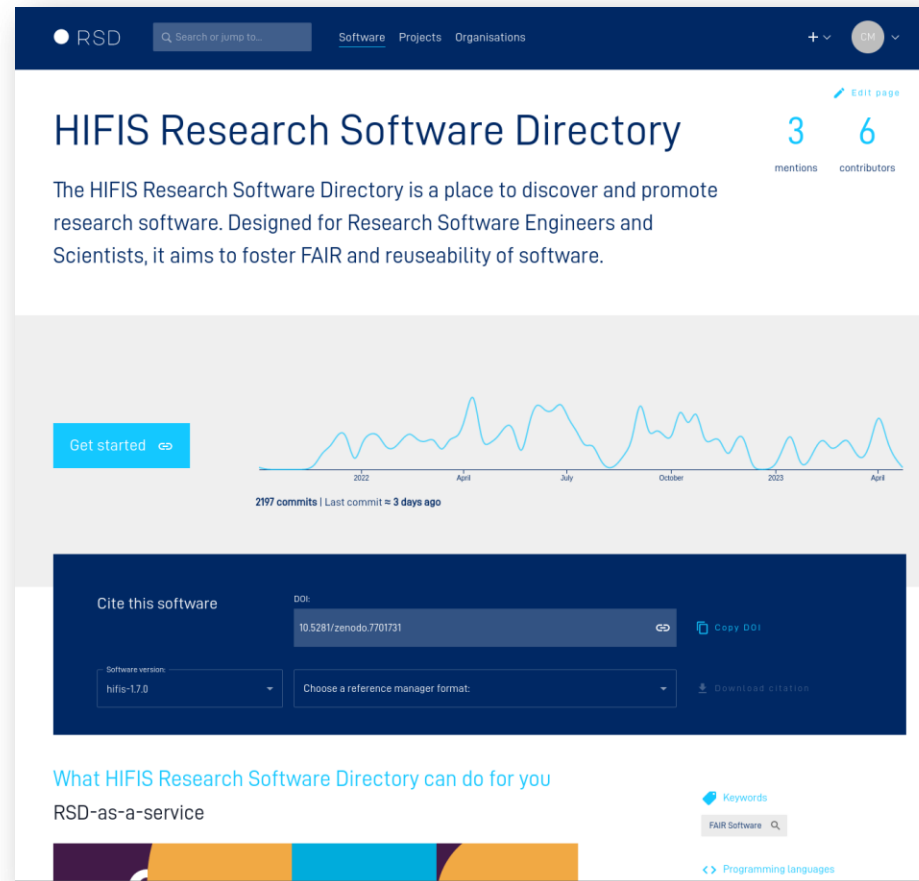
Contributors

References

The HERMES Workflow
…with Research Software Directory (RSD) Extension

DMA-8: Integration of near-real-time/online data analysis solutions for extreme scale data into the software toolbox of DMA

The software tools developed at the GSI within the DMA ST2 Activities has been used to deploy and control ~70000 tasks on 2200 GPUs and about 20000 CPUs in the online farm directly connected to the ALICE detector. Slurm plugin developed within the ODC software allows deploying on any Slurm-controlled clusters to be used later by the GSI/FAIR experiments and MT-DMA communities with similar huge demands on data rates

ALICE
Pb-Pb 5.36 TeV
LHC22s period
18th November 2022
16:52:47.893

DRESDEN concept

HZDR
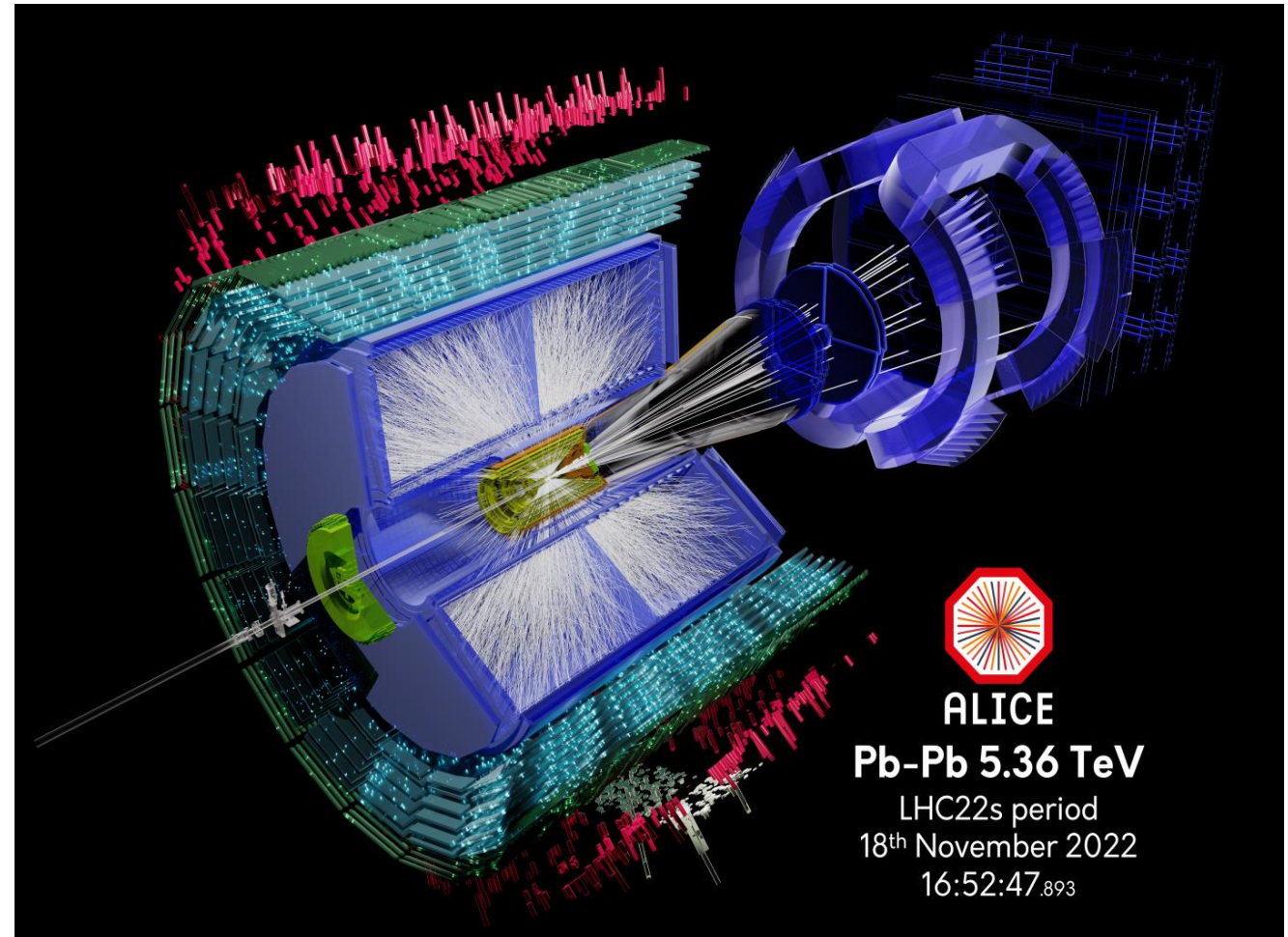
Data-parallel Types for C++

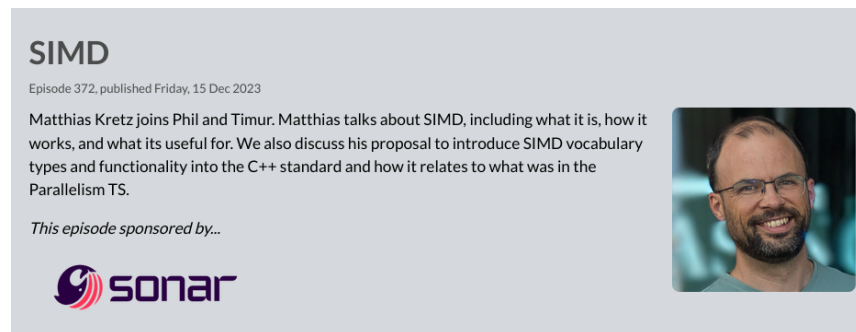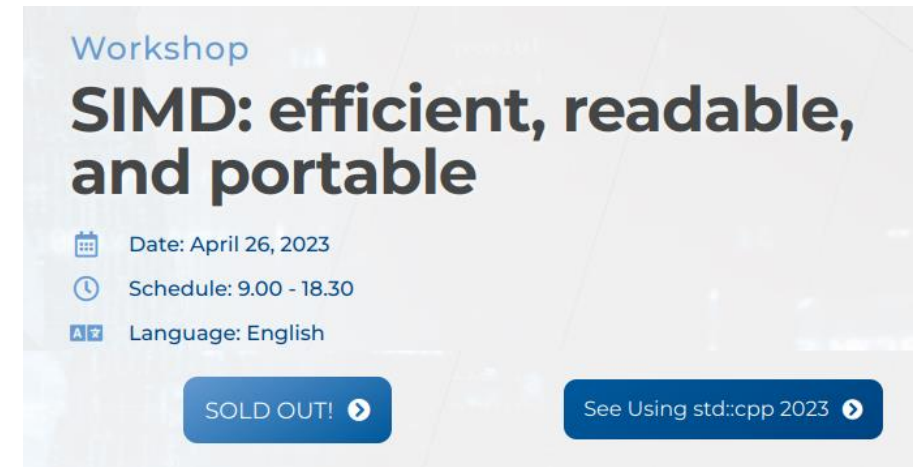GCC implementation fully usable and documented on cppreference.com

Tutorial on std::experimental::simd at "using std::cpp 2023"

Design approval of std::simd for C++26

Collaboration with Intel on adding more features to simd in C++26

CppCon 2023 presentation

CppCast interview

Exchanging tools and methods on all levels

Lessons learned

Basic software engineering

- Abstracting fast changing/varying hardware components
  - Heterogeneous hardware (e.g. Alpaka)
  - SIMD (Vc → lead to std::simd, now part of C++)
- Committee/Community work
  - C++
  - GCC
  - OpenACC
  - OpenPMD
- Automate testing/packaging
  - Use HIFIS https://codebase.helmholtz.cloud CI resources to test everything
  - Even HPC integration now (locally) possible (Jacamar), soon also in Helmholtz Cloud

Research Software Engineering is the new community term

Building even more bridges

DMA creates/contributes to software on all levels

Research software infrastructure

1 It involves research software that captures more broadly accepted and used ideas, methods and models for use in research, and warrants close researcher involvement in their development.

Prototype tools

2 It refers to research software that demonstrates a new idea, method or model for use by others outside the project within which it originated, often as a substantive intellectual contribution in its own right and often in the form of a proof of concept.

Analysis code

3 It includes research software that captures computational research processes and methodology, and often occurs in the context of simulation, data generation, preparation, analysis and visualisation.

abundance

Importance

Foundational Software

ESCAPE – Latest Developments: Collaboration

Punch4NFDI LoI – OSSR as one area of collaboration:

27 September 2024

ESCAPE Open Collaboration and PUNCH4NFDI Consortium sign landmark Letter of Intent

The ESCAPE Open Collaboration and the PUNCH4NFDI Consortium announce the formal signing of a Letter of Intent.

READ MORE

The road ahead

Transitioning from POF IV to POF V

Strengthening the network

- DMA is not a team, but a loose network of strong pont-to-point connections between highly demanded groups
- Key stakeholders are also in the network beyond direct DMA affiliations

AI models very successful

- Not yet well visible in RSD
- Needs maybe other outlet to integrate?

Avoid too much micro-planning

- DMA does great work --> worry about labeling later

DRESDEN concept

HZDR