



Services Roadshow by



Dec 04, 2024

M. Hammitzsch et al.

Challenges



- Find research software
- Understand its connection to other research artefacts such as data
- Evaluate and reuse software for specific research questions and for one's own research
- Ensure the sustainable use and further development of research software
- Expand the spectrum of analysis and processing options for research data
- Establish a cross-consortia initiative focused on common objectives since much work regarding research software is planned and conducted within at least 20 consortia

Goals

- Serve single entry point with information about research software within the NFDI
- Improve accessibility and use of research data
- Integrate different scientific communities
- Link and coordinate independent individual developments
- Extract, standardise and enrich research software metadata
- Provide the foundation for establishing standards and for networking across the NFDI
- Minimise duplication of work, maximise compatibility to similar and complementary efforts



Goals of the initialisation phase

The nfdi.software proof-of-concept service includes a website and an API, providing an interface to search and discover research software via the NFDI. Identification of needs and community specific requirements will be collected in collaboration with consortia and basic services through workshops, discussions and prototyping. The 'Prototype for Integration' will feature functionalities for harvesting and retrieving information, offering detailed software data along with related publications, people, and projects. Existing tools like the [Research Software Directory](#), [bio.tools](#), the [Research Software Ecosystem](#), [Betty's Research Engine](#) and [physics.tools](#) will be integrated, and an advanced search and recommender system will be added. Various metadata vocabularies will be explored and discussed to develop a common cross-sector vocabulary. Metadata efforts like Knowledge Graphs and Jupyter4NFDI will be evaluated, with results documented in a concept and design report for feedback and improvement. The integration with I4M4NFDI is also planned to foster synergies.

Project plan



The Research Software Directory promotes the impact, re-use and citation of research software.

base4
nfdi

DFG Deutsche
Forschungsgemeinschaft

Cookies
User Documentation
Technical Documentation

Approach

Design, implement and test a central research software marketplace

- Bring together players and build on top of their **existing solutions**
- Connect **complementary components**
- Provide a tangible starting point for **first-hand experiences**
- Offer an early **proof-of-concept service**
- Map and link distributed metadata
- **Link base services**
- Enable user feedback
- Identify strengths and **weaknesses, problems and gaps**
- Update iteratively towards a prototype service



Betty's (Re) Search Engine



physics.tools

 TS4.nfdi

 KGI4.nfdi

 IAM4.nfdi

 Jupyter®4.nfdi

 PID4.nfdi

 DMP4.nfdi

Approach

Guide the future development in a follow-up integration phase

- Establish communication for a **concerted effort**
- Establish bridges to the **domain-specific NFDI infrastructures and services**, and the corresponding **communities** with a focus on research software
- Bring together reliable contributions **from consortia** with the necessary impetus, at present NFDI4Earth, DataPLANT, NFDI4BioImaging, NFDI4Ing, PUNCH4NFDI, Text+, NFDI4Culture, NFDI4Health, NFDI4Energy, NFDI4DataScience, NFDI4Cat and others
- Jointly design the **interplay of related consortia services**
- Incorporate experiences and results of other initiatives



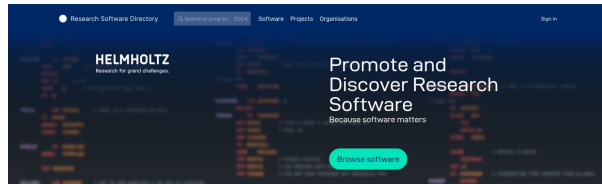
Components of nfdi.software

Build on proven solutions - establish standards - guarantee compatibility

1. **Research Software Directory (RSD)**
aims at finding and reusing cross-domain research software, stimulating reuse, encouraging proper citation, and making the impact of research software more visible
2. **ELIXIR Research Software Ecosystem (RSEc & bio.tools)**
supported by the European Infrastructure for Biological Information (ELIXIR), where researchers can find and compare bioinformatics tools thanks to curated metadata
3. **Betty's Research Engine**
search engine that finds software repositories and links them to corresponding scientific publications and enriches the results with harmonised metadata
4. **Physics.tools**
searches in publications for referenced software and automatically retrieves accessible metadata information from the code repository
5. **LLM based search engine**
allows to find publications and related software products without needing to know the name of a given software, allows find multiple, similar products

1. Building Block Research Software Directory

A home for research software



Latest news

44846

Mentions of research software in science

14 NOVEMBER 2024
Helmholtz RSD updated

Helmholtz Program-oriented Funding IV

13 SEPTEMBER 2024
Program-oriented Funding in the Helmholtz RSD

Helmholtz Software Award

3 SEPTEMBER 2024
Helmholtz Software Award Call 2024

News archive

The Helmholtz RSD in numbers

339

Registered software packages

1713

Contributors to research software

44970

Mentions of research software in science

197

International partner organisations

Software Spotlights

The latest outstanding software product developed in Helmholtz.



Kadi4Mat

Kadi4Mat

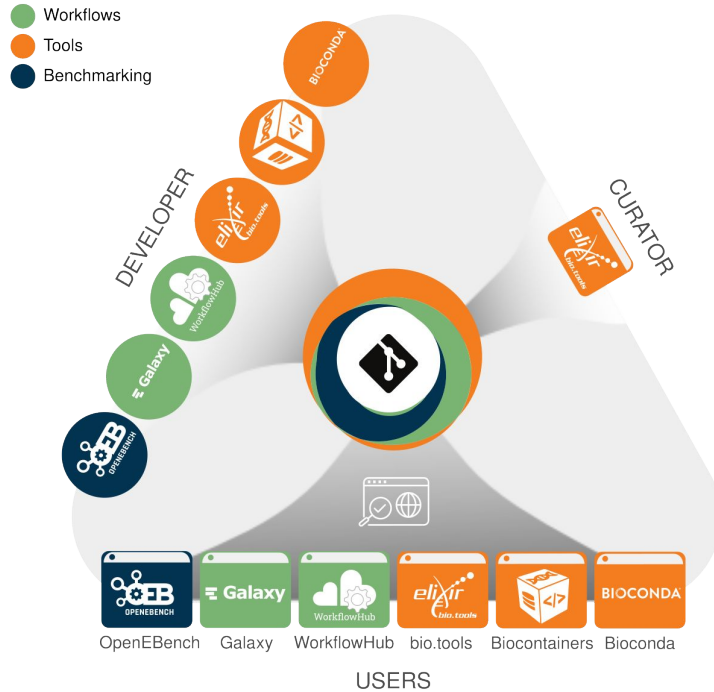
Kadi4Mat is a generic and open source virtual research environment.

Browse all Spotlights

- A digital marketplace for research software
- Dedicated to improve findability, citeability and impact
- Relates research software to
 - Other research software
 - Research projects
 - Organisations
 - Contributors
- FOSS, originally maintained by Netherlands eScience Centre (NLeSC), Helmholtz joined development
- Currently productive instances
 - <https://research-software-directory.org> by NLeSC
 - <https://helmholtz.software> by Helmholtz Federated IT Services

2. Building Block *bio.tools*

ELIXIR Research Software Ecosystem (RSEc)

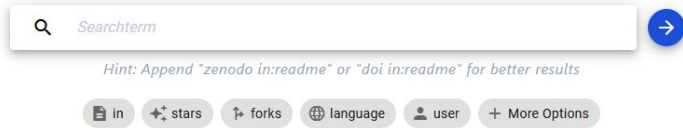


- Centralized and curated tool metadata repository
 - Automatically collects metadata from various sources
 - Galaxy metadata includes usage statistics and tool availability
 - Allows for the synchronization between different sources
 - The repository can be used as base to develop frontend solutions
 - <https://bio.tools/>

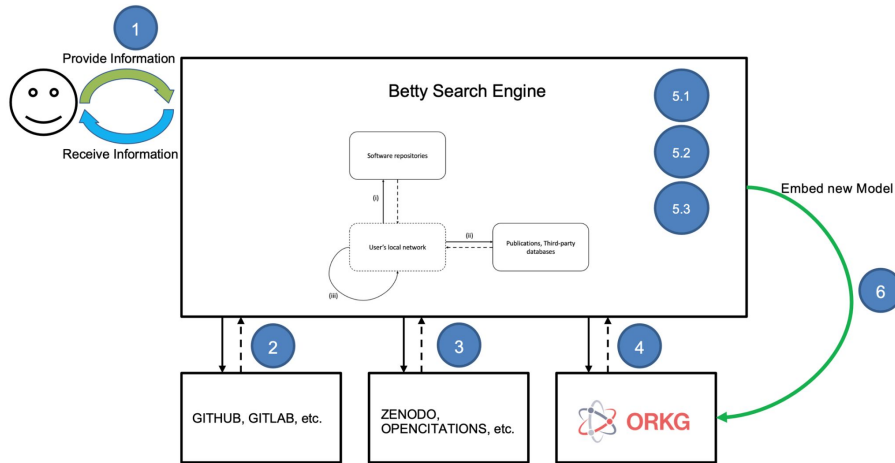
3. Building Block Betty's Research Engine



Betty's (Re) Search Engine



LINK TO OUR PAPER LINK TO OUR GITLAB VIDEOS CONNECTION STATUS



- Meta Search Engine for Research Software

- Cascadic Search for:

- SW Repositories
- Corresponding Publications
- Knowledge (from ORKG)

- Extendable Architecture

- Decentralized, no central Server!

- <https://nfdi4ing.rz-housing.tu-clausthal.de>

4. Building Block *physics.tools*

The screenshot displays the physics.tools search interface. At the top left is the physics.tools logo. To its right is a dropdown menu set to 'ALL', a search input field, and a 'Search' button. Below the search bar, the results for 'gammapy' are shown. A 'Github repository' card is prominent, listing details such as the link to the repository, a description ('A Python package for gamma-ray astronomy'), language (Python), license (BSD 3-Clause), and statistics like 30 contributors, 197 forks, 231 stars, and 22 watchers. Below this, a 'Citations in arxiv publications' card features a line graph with a green line showing an upward trend. At the bottom, a 'Publications mentioning this software:' card lists three articles with their titles, dates, and associated categories.

Field	Value
Link	https://github.com/gammapy/gammapy
Description	A Python package for gamma-ray astronomy
Language	Python
License	BSD 3-Clause "New" or "Revised" License
Contributors	30
Forks	197
Stars	231
Watchers	22
Open Issues	231
Created	2013-05-15T07:50:...
Updated	2024-11-12T17:38:3...

Publications mentioning this software:

- How Do Software Developers Use GitHub Actions to Automate Their Workflows?**
2021-03-22 | pdf | Computer Science - Software Engineering
- Physics Performance of the Large-Sized Telescope prototype of the Cherenkov Telescope Array**
2021-09-08 | pdf | physics.astro-ph | Astrophysics - High Energy Astrophysical Phenomena | Astrophysics - Instrumentation and Methods for Astrophysics
- The Future of Gamma-Ray Experiments in the MeV-EeV Range**
2022-03-14 | pdf | physics.astro-ph | Astrophysics - High Energy Astrophysical Phenomena | Astrophysics - Instrumentation and Methods for Astrophysics
- Provenance of astronomical data**
2022-04-25 | pdf | physics.astro-ph | Astrophysics - Instrumentation and Methods for Astrophysics | Astrophysics - High Energy Astrophysical Phenomena

- Software search engine
 - Database filled from software links found in arxiv publications
 - Provides statistics on software products (e.g. citation history, usage statistics from host platform, etc.)
 - Shows all papers with the same link
 - Shows additional software links from these papers
 - <https://physics.tools/>

5. Building Block *LLM based search engine*

- Fine-tune a Large Language Model already trained for general purposes, e.g. a [GPT4ALL model from Nomic-AI](#)

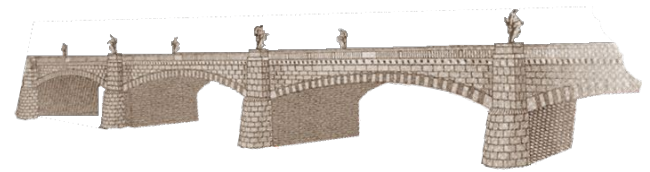


- Download the model, all the weights and the pretrained parameters
- Fine-tune it on our dataset, the database of arxiv publications:
 - Parameter-Efficient Fine-tuning (PEFT) on a small number of extra parameters
 - Low-Rank Adaptation (LoRA), freeze pretrained model weights
- Forecasted use case:
 - Extend the physics.tools search engine capabilities with scientific Q&A on the abstracts and bodies of the articles in the database

Research Software Metadata

Semantically structured metadata for Research Software

- Why
 - Key to implement FAIR for Research Software
 - Bridge towards other research artifacts
- How
 - Analysis of existing approaches
 - Understanding NDFI specific needs
 - Recommendation to reuse and tailor it to consortia



CodeMeta



Bioschemas

Community & Networking

Securing community acceptance

- Addressing the NFDI community
 - Review of already existing solutions, plans and policies
 - Stakeholder workshops
 - Prototype presentations
 - Survey for needs
- Monitor international developments in RDA and EOSC
- From fit-for-use to fit-for-purpose
 - Collecting further needs and requirements from NFDI
 - Develop proposals for curation and quality criteria



Work Programme

Initialisation phase, Nov 2024 - Oct 2025

Project month	1	2	3	4	5	6	7	8	9	10	11	12
WP1 Prototype Service		1.1	1.2		1.3	1.4				1.7		1.8
					1.5	1.6						
WP2 Community	2.1		2.2			2.3	2.4			2.5		2.6
WP3 Concept & Design			3.1			3.2				3.4		3.5
WP4 QA & Coordination	C1		C2			C3	C4			C5		C6

- **WP1 Prototype Service**
 - Covers the implementation and initial integration of complementary components and services
 - Provides proof-of-concept first and, later on, a continuously updated prototype service
- **WP2 Community Building & Networking**
 - Ensures community-wise interactions and NFDI stakeholder involvement
 - Nurtures an active exchange towards a concerted effort within the NFDI
- **WP3 Concept & Design**
 - Feeds from the prototype experiences and feedback as well as the stakeholder exchange as well as the expressed community needs
 - Drafts design and concept for integration in a follow-up project
- **WP4 Coordination and Quality Assurance**
 - Takes care of coordination and management
 - Ensures delivery of expected results and quality

Call to Action

Deliverables depend on contributions & participation

- WP2 Community Building & Networking
 - D2.2 **Descriptive review of already existing solutions**, plans and policies regarding software marketplaces in NFDI consortias → [January/February](#)
 - D2.3 **Survey for needs** and current implementations in NFDI → [April](#)
 - D2.4 Public **interim presentation and workshop** with consortia and base services providers → [May/June](#)
- WP3 Concept & Design
 - D3.1 Identification of relevant metadata recommendations and vocabularies for research software and **crosswalks to NFDI consortia marketplaces** and registries, in cooperation with NFDI-Meta RSMeta WG → [January/February](#)
 - D3.2 Initial concept for **metadata exchange with other Base4NFDI services** → [April](#)
 - D3.5 Concept and **design for successive integration phase** → ~~October~~ [April or July](#)

What will nfdi.software offer?

- PI)** a hub that develops software for NFDI on demand
- OM)** an NFDI RDM service portfolio
- DA)** a central marketplace for NFDI research software



Thank you!
Questions?



-  nfdi-software@lists.nfdi.de
-  base4nfdi-servicestewards@lists.nfdi.de for general inquiries
-  **TBA**