nfdi Nationale Forschungsdaten Infrastruktur

Jupyter4NFDI
A central JupyterHub for the NFDI

Björn Hagemeier
2024-07-15

Overview

- About Jupyter

- History

- Current project

- Outlook

- Jupyter's mission is to support *reproducible, exploratory and interactive computing* through
  - A collection of standards
  - A community
  - A set of software tools
- Concepts
  - Jupyter *Notebook*: Combination of documentation & programming code
  - *JupyterLab*: Collection of Jupyter Notebooks and data
  - *JupyterHub*: Multi-user, web-based Jupyter environment. Avoids the need to install software, allows for sharing. Facade for JupyterLab, but also RStudio and others
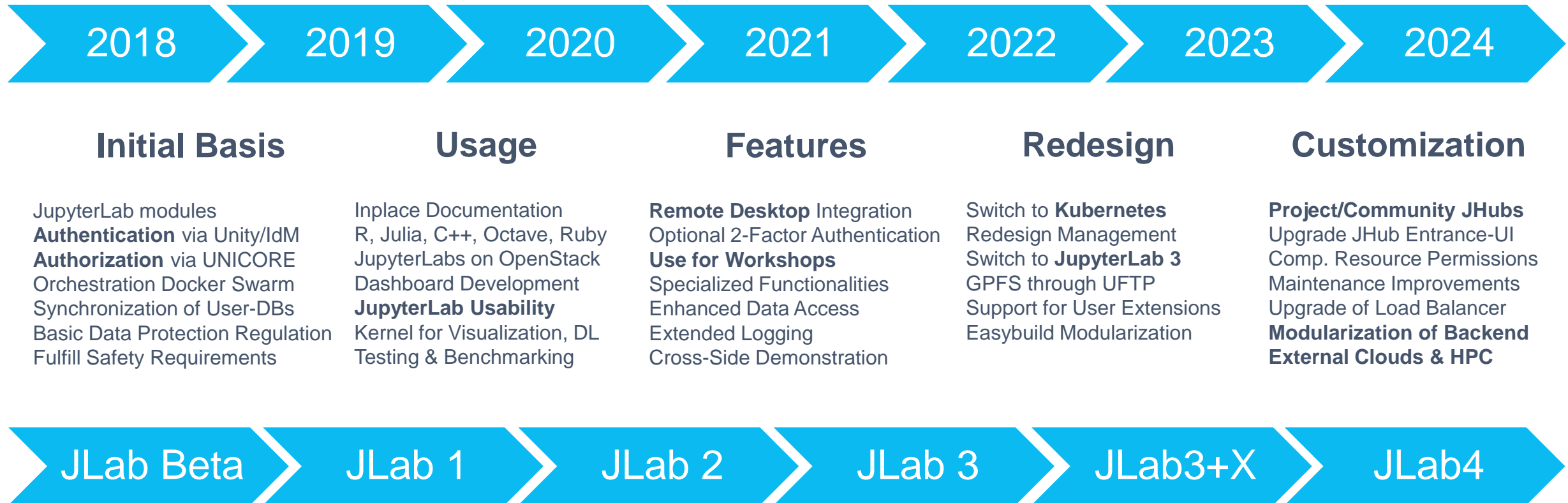
Jupyter in NFDI consortia

- 6 of 26 (23%) reported comparable service **in use**.

- 17 of 26 (65%) reported a **need for** Jupyter or comparable service.

- 8 of 26 (31%) reported to be **developing** a comparable service.

- Numerous **constructive comments and questions raised** during feedback round. We take them into consideration.
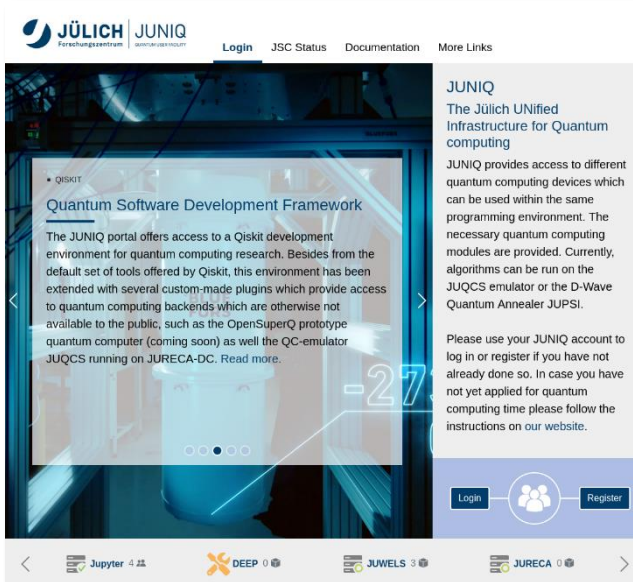
Jupyter at JSC

- Providing a highly flexible Jupyter Hub facilitating access to **HPC and Cloud** resources since 2017.

- Deployment on very stable **Kubernetes** cluster inside Cloud environment.

- **OutpostSpawner** to integrate external resources.

History of Jupyter at JSC

2018	2019	2020	2021	2022	2023	2024

Initial Basis	**Usage**	**Features**	**Redesign**		**Customization**
JupyterLab modules	Inplace Documentation	**Remote Desktop** Integration	Switch to **Kubernetes**		**Project/Community JHubs**
Authentication via Unity/IdM	R, Julia, C++, Octave, Ruby	Optional 2-Factor Authentication	Redesign Management		Upgrade JHub Entrance-UI
Authorization via UNICORE	JupyterLabs on OpenStack	**Use for Workshops**	Switch to **JupyterLab 3**		Comp. Resource Permissions
Orchestration Docker Swarm	Dashboard Development	Specialized Functionalities	GPFS through UFTP		Maintenance Improvements
Synchronization of User-DBs	**JupyterLab Usability**	Enhanced Data Access	Support for User Extensions		Upgrade of Load Balancer
Basic Data Protection Regulation	Kernel for Visualization, DL	Extended Logging	Easybuild Modularization		**Modularization of Backend**
Fulfill Safety Requirements	Testing & Benchmarking	Cross-Side Demonstration			**External Clouds & HPC**

JLab Beta	JLab 1	JLab 2	JLab 3	JLab3+X	JLab4

Jupyter Hub infrastructure at JSC
A hosting environment for several hubs

- CoE in combustion

- CoE RAISE – AI for simulation base engineering

- EuroCC-GCS

- GCS Portal

- Juniq – Quantum Computing in your browser

- Jupyter-JSC

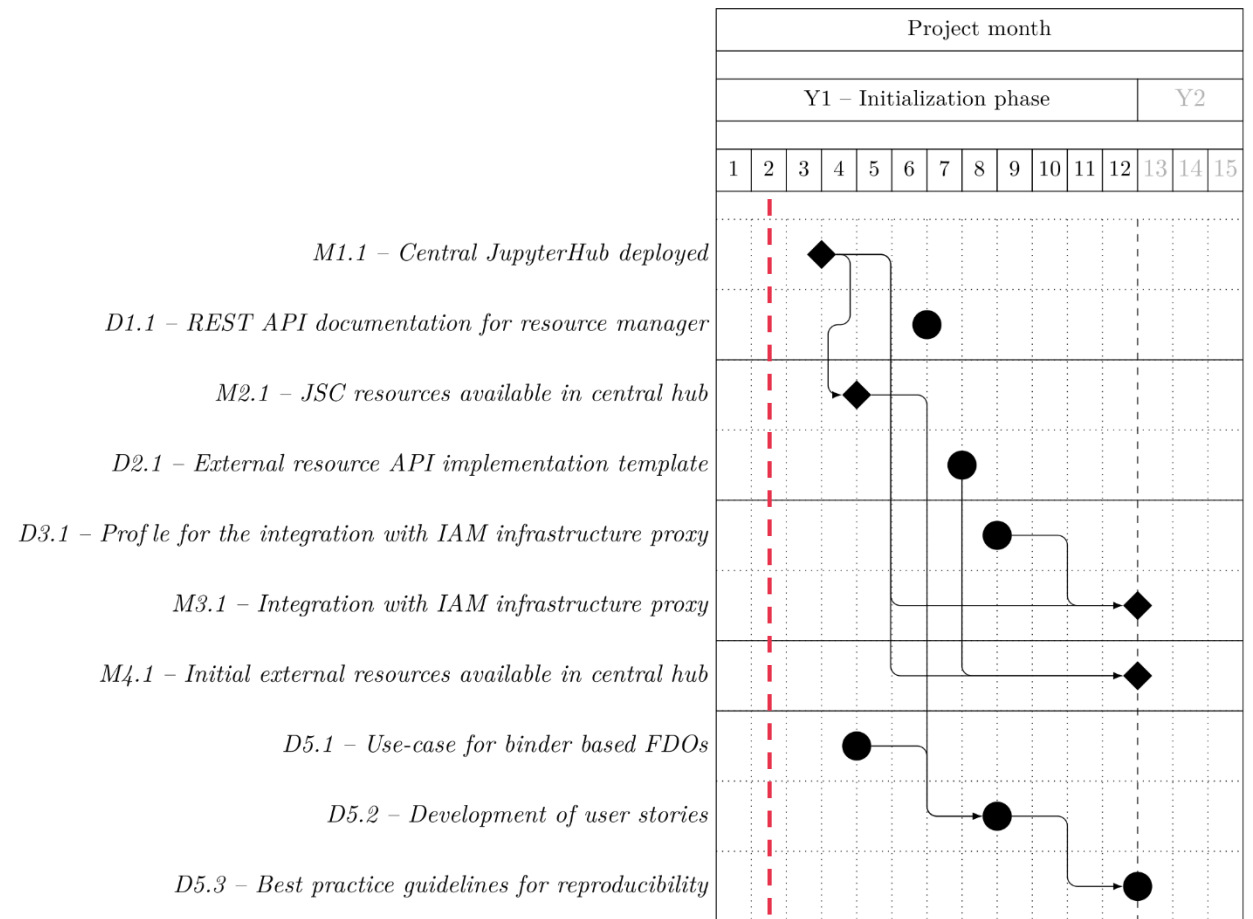- **Jupyter4NFDI**

RSE within section common infrastructures

- Meetings of WG RSE focusing on Jupyter as early as June 2022

- Invited core developers of upstream Jupyter who followed with interest

- Preparation of proposal for base service between February and August 2023

- Successful on second submission February 2024

- Start of project in June 2024

- Reached M1.1 and M2.1 this month

  - Starting point for further developments

  - Integration of additional resources

    pending (M4.1)

- User stories (D5.2) in collaboration with

  interested parties and supported by

  Base4NFDI

Jupyter4NFDI
Participants and their roles

- CIDS: NHR resource provider
- GESIS: user representation, outreach
- GWDG: IAM integration and resource provider
- JSC: lead, technology and resource provider
- MPCDF: resource provider
- US: WP4, connecting external resources, link to datasets

Connecting external resources

- Inventory of Jupyter services within NFDI
- https://nfdi-jupyter.de
- Some will be immediately integrated
- Some guarded by additional security restrictions
- Others welcome to join

Relation to other basic services

IAM4NFDI

- Authentication at basic service
- Extraction of attributes to tailor UI according to user permissions
- Pre-selection of resources
- Restriction of available options

Multi-Cloud

- Integrate as external resource at multiple locations of the Multi-Cloud
- Potential hosting environment for central service

PID4nfdi

- References to data and software (including executable notebooks)
- Use of PIDs to reference FAIR Digital Objects (FDO)
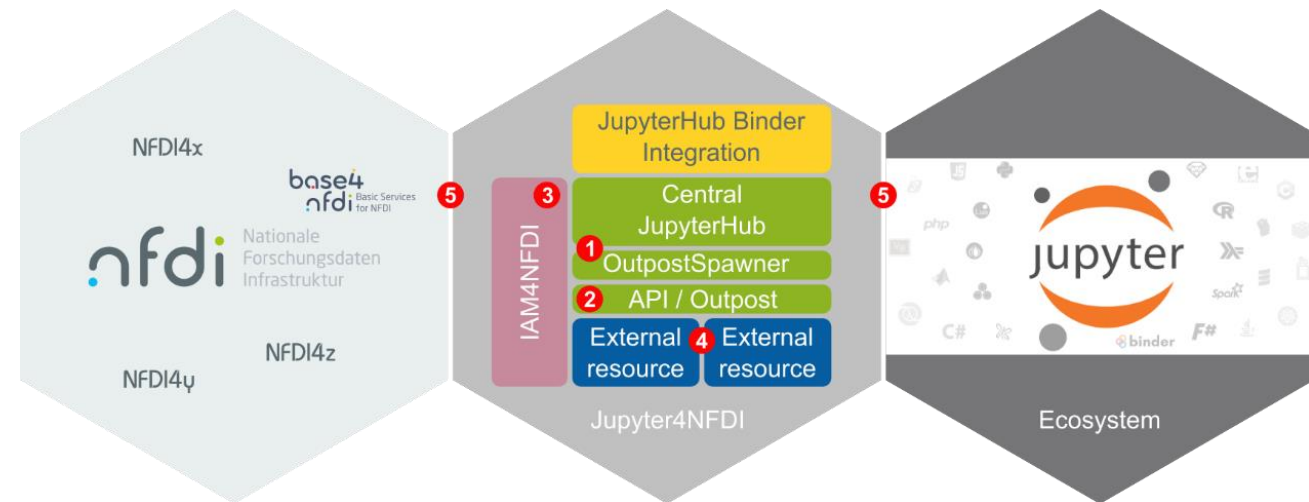
Community support

Platform for a comprehensive toolbox for sheet music analysis and visualization

...

Development of RDC. Integration of existing, Jupyter-based toolbox in Jupyter4NFDI

Integration of FEL software stacks with HPC systems

Available now

Initial Jupyter Hub for NFDI

- Minimal Cloud-based resources

- The more RAM and CPUs, the less runtime

- Custom images

- Configurable Extensions and Kernels

- Visit https://hub.nfdi-jupyter.de/

Demo

- Provide a sound, sustainable and user-friendly infrastructure, scaling to the national level

- Support more consortia within NFDI
  - Long-tail users with generic resources
  - Users with increased and very specific resource demands

- Data management and availability

- Ample training material about, as well as community trainings conducted on **Jupyter4NFDI**