

USER's View: Evolving User Demand

- interactive environment, GUI
- large to huge datasets
- large compute resources (required CPUs, RAM, GPU)

ADMIN's View: Efficient Infrastructure Usage

- expensive hardware, short lifespan
- batch jobs with better efficiency than cloud usage
- better resource planning with batch jobs (PBS)

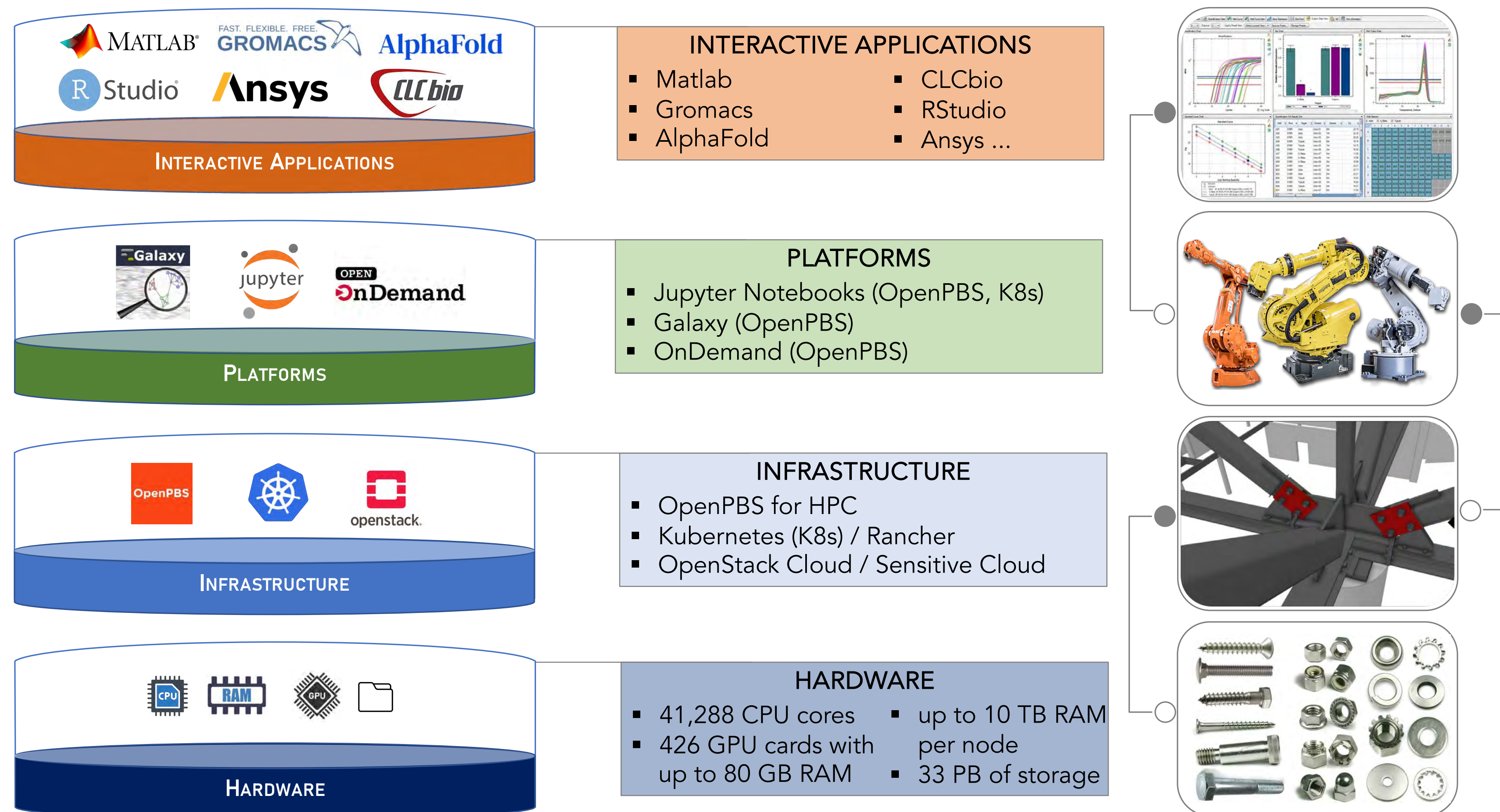
Interactive vs. Non-interactive

For decades, resources provided by large computing and data infrastructures have been accessed by **non-interactive batch job submission** and data storage on shared filesystems.

There are good reasons to retain this approach, as it still offers optimal performance and the **best utilization** of precious resources. On the other hand, it suffers from **limited user comfort** and support for recently emerging requirements of reproducible science.

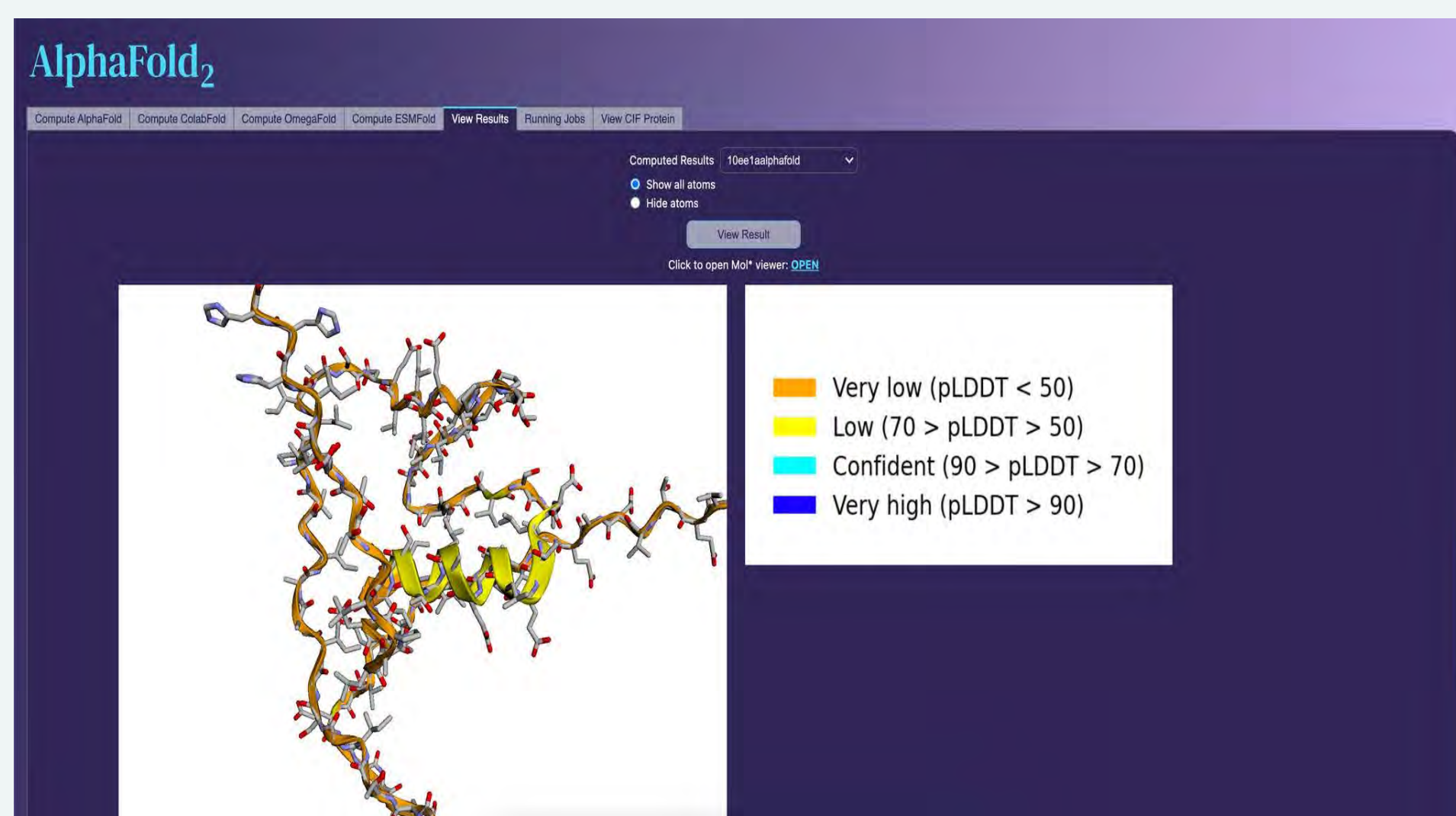
Therefore, e-INFRA CZ provides and fully supports alternative computing **environments for interactive work**.

<https://metavo.metacentrum.cz/>
<https://docs.metacentrum.cz/>
<https://docs.cerit.io/>



Usecase1: Alphafold and more as Jupyter Apps

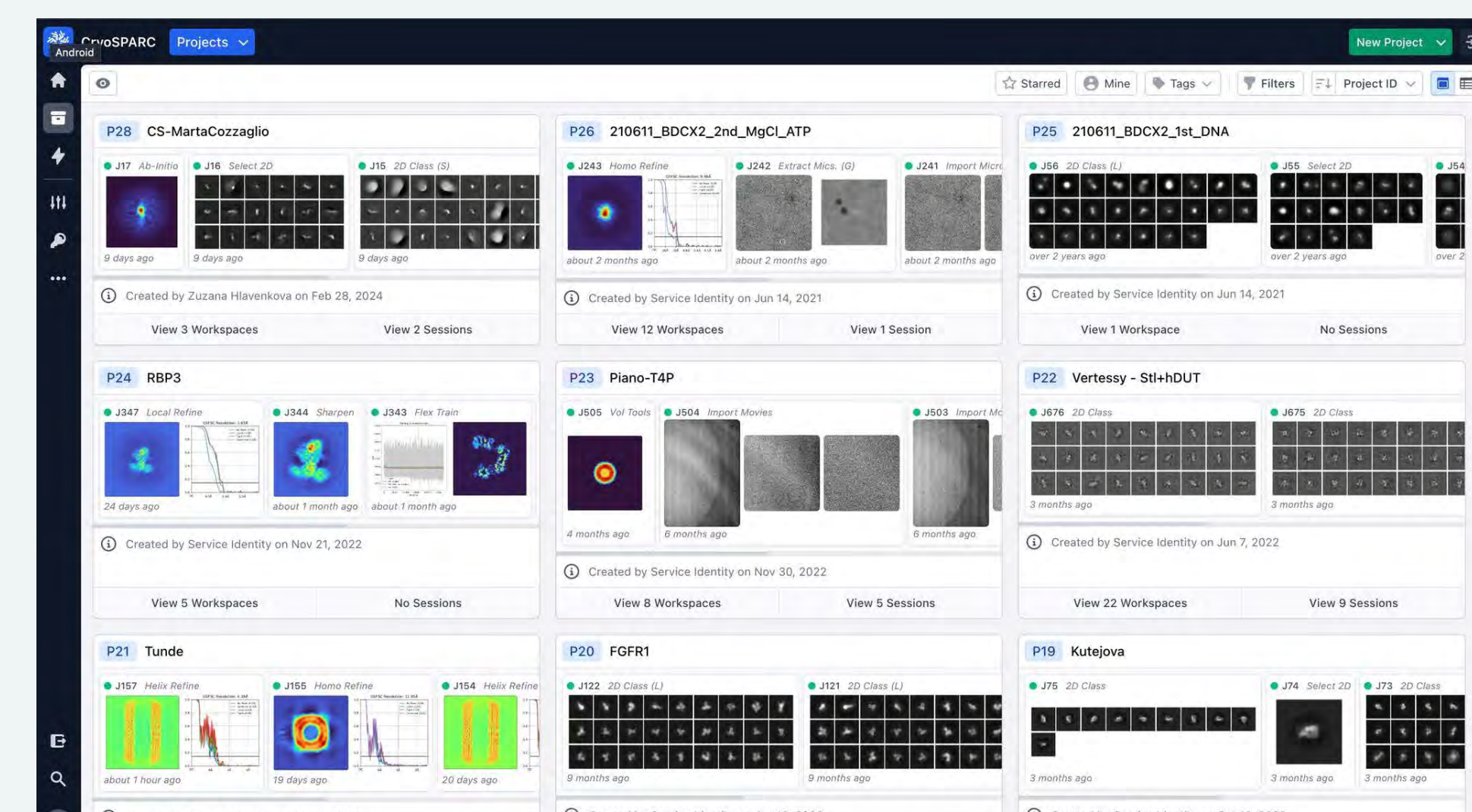
With the use of widgets libraries, experimental **Jupyter notebooks** can evolve continuously into **full-featured GUIs**. Currently, we provide such interfaces for Alphafold / Omegafold / Esmfold and a pilot for the molecular dynamics hub; others may emerge according to the community requirements.



<https://alphafold.cloud.e-infra.cz/hub/>

Usecase2: CryoSPARC in Kubernetes

Multiple parallel instances of CryoSPARC can access **large GPUs** (NVIDIA A100 with 80 GB RAM), and **submit** computations to both **PBS** and **Kubernetes** while still providing the interactive experience. CryoSPARC, an essential tool for structural biology, enables the extraction of valuable information from single-particle cryo-EM data.



About e-Infra CZ

The main aim of e-INFRA CZ is to provide a unique e-infrastructure for research in the Czech Republic. It offers computing and storage capacities and related services for transmitting, storing, and processing scientific data, enabling complex simulations, modelling, and data analysis.

The National Repository Platform, a core component of the National Data Infrastructure (NDI), is being developed as part of this endeavour.

<https://www.e-infra.cz/>

Forever FAIR

Many of the interactive platforms (Jupyter notebooks and Galaxy in particular) address the **reproducibility** requirement ("R" from FAIR) natively by recording the history of user's calculations, as well as **interoperability** ("I") to some extent (support of data format conversions, etc.).

e-INFRA CZ gradually develops tools to address **findability** ("F", e.g. pilot of semi-automated metadata provisioning for molecular dynamics calculations). Data FAIRness altogether, including **Accessibility** ("A"), is the principal goal of the current development of the **National Repository Platform** (NRP).

NRP will provide **approx. 50 PB** of usable storage capacity, **core services** (Invenio, DSpace), **catch-all** repository, and technical support. User communities are required to **contribute** (define repository-specific metadata, policies, and processes) via **EOSC working groups**.

<https://www.eosc.cz/>
<https://data.narodni-repozitar.cz/>



<https://ces.net/NH2024>

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

MUNI CERIT-SC

cesnet

cesnet metacentrum