Imaging Data in EOSC

COVID-19 as Demonstrator

Science Project

Reagon Karki (EU-Openscreen/Fraunhofer ITMP)
PhD in Bioinformatics

EOSC Symposium, Nov 14-17, 2022

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020, Grant Agreement 101017536

The Scientific case in a nutshell: CONTEXT

COVID-19 Outbreak

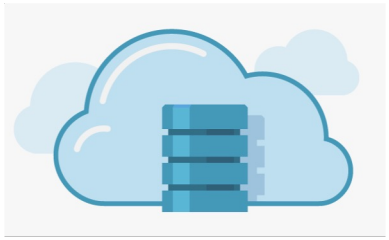
Imaging Community Needs

- **Efficient and open image data sharing** among researchers and their communities
- **Re-usability** of the vast amount of imaging data generated to accelerate the response.
- **Data harmonization** and **interoperability** to achieve a better understanding of the disease and molecular basis.

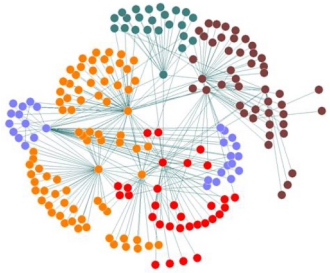
Imaging Community Issues

- Difficulties accessing **big, high-dimensional image data** located in the cloud storage
- Lack of **standardisation** for **data and metadata** to make images and analyses reusables
- Lack of **harmonization** between **high-level imaging data** with **structural data** that helps extending its molecular context.

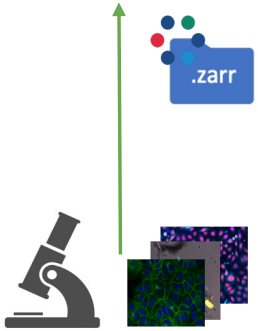
The Scientific case in a nutshell: TSP Action



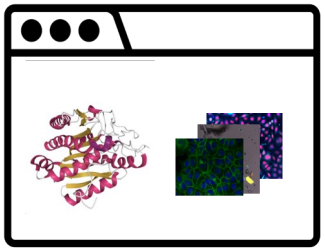
Public repositories



COVID-19 Knowledge Graph



Bioimage Acquisition & Analysis



3DBionotes-WS +
the COVID-19 Structural Hub

EU-OS hit
compounds

TSP Objectives

1) Implement a community standard cloud-compatible open image data format and data submission workflow

1) User-friendly access to multidimensional structural image data and quality-related metadata in EOSC space

1) Analysis pipeline for large scale COVID-19 functional and image based data sets for integration with complementary resources

1) Knowledge graph deployment embedding COVID-19 phenotype-chemotype and integration of project partner data

EURO BIOIMAGING

instruct ERIC

eu-openscreen

Fraunhofer ITMP

Scientific case

Technical requirements

- integration and interoperability of research data and services among imaging domains
- software/platforms sustainability over time
- user-friendly and centralized way of accessing new platforms within the imaging domain
- Deployment of resources
    - EuroBioimaging workflow
    - Fraunhofer ITMP KG

EOSC as enabler (added value)

- Centralized place to easily find available platforms for the user needs (EOSC Marketplace)
- Facilitate platforms usability for providers to check their impact on the community (EOSC Core Accounting)
- Support and unified way of communication between users and platform developers (EOSC Core Helpdesk)
- Gain insights into anomalous behavior for platform providers to achieve a good performance (EOSC Core Monitoring).
- Centralize and unify data exploiting and sharing among the different communities (EOSC Horizontal Services).

Scientific platform – Integration with EOSC

Achieved integrations

Core Services:

- Monitoring for 3D-Bionotes-WS (ongoing)
- Helpdesk for 3D-Bionotes-WS (ongoing)

Next steps

Core Services:

- Accounting for services for 3D-Bionotes-WS.
- Accounting for services, Monitoring and Helpdesk for the COVID-19 Structural Hub

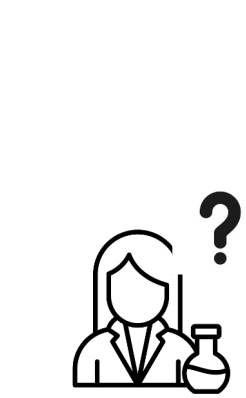
Horizontal Services:

- Galaxy workflow for image data conversion to OME-Zarr and submission to Bioimage Archive is to be deployed to WorkflowHub
- COVID-19 KG (https://github.com/Fraunhofer-ITMP/BY-COVID-KG)

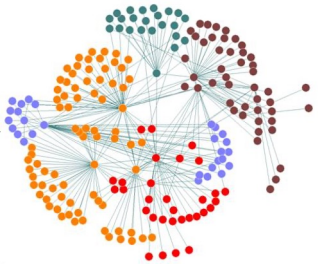
Impact

- Enable **publication, access, sharing and analysis** of imaging data and metadata since the COVID-19 outbreak.
- **Well-established infrastructures** to act better and faster in **future pandemics**.

The demonstration – step-by-step



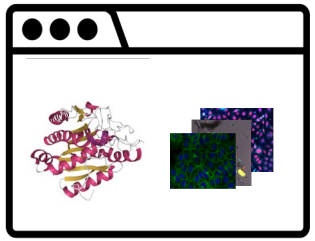
Landscape of available compounds to find ways to treat Monkeypox?



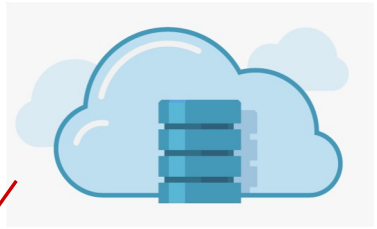
COVID-19 Graph

Knowledge

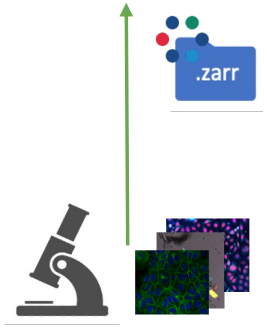
EU-OS hit compounds



3DBionotes-WS + the COVID-19 Structural Hub



Public repositories

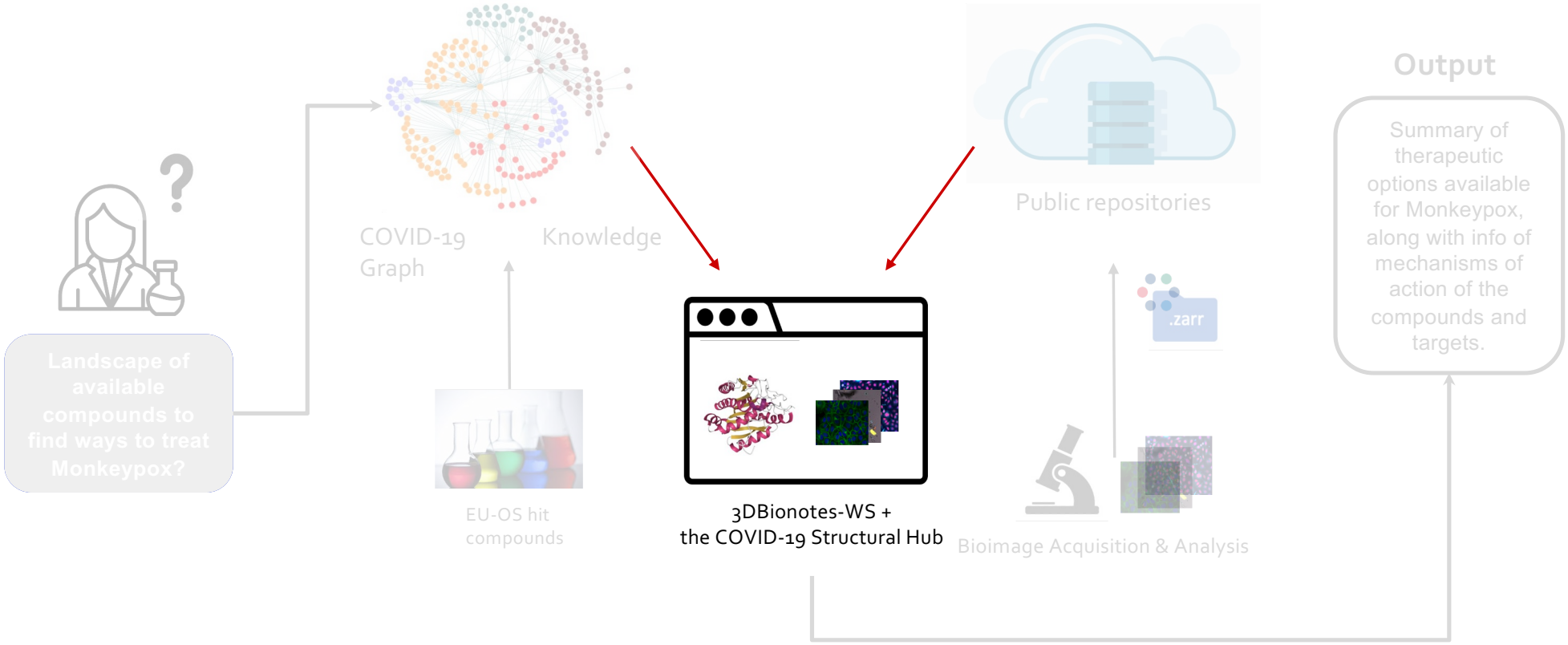


Bioimage Acquisition & Analysis

Output

Summary of therapeutic options available for **Monkeypox**, along with info of **mechanisms of action** of the **compounds** and **targets**.

The demonstration – 3DBionotes-WS workflow



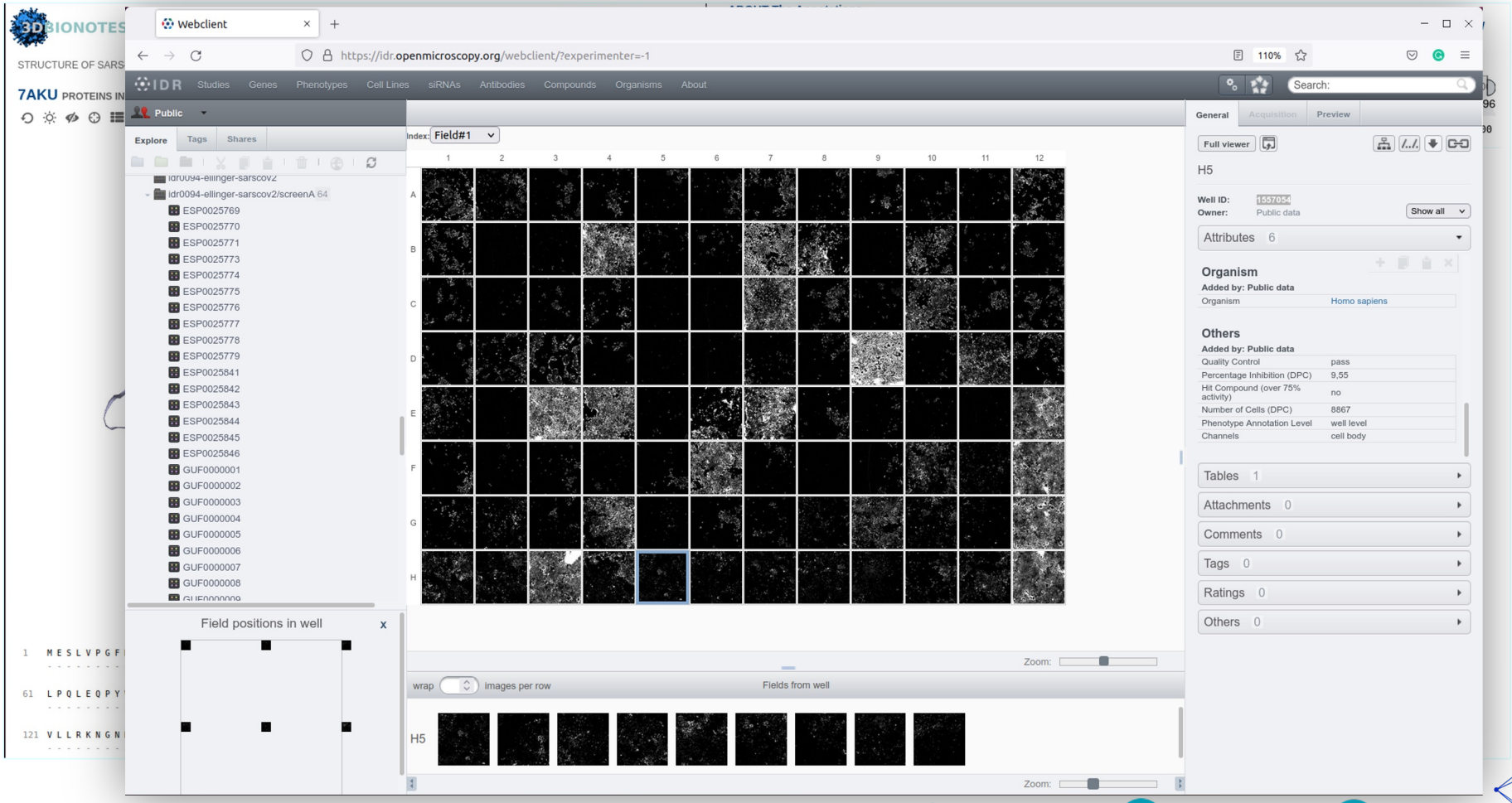
The demonstration – 3DBionotes-WS workflow

In production

The demonstration – 3DBionotes-WS workflow

In development

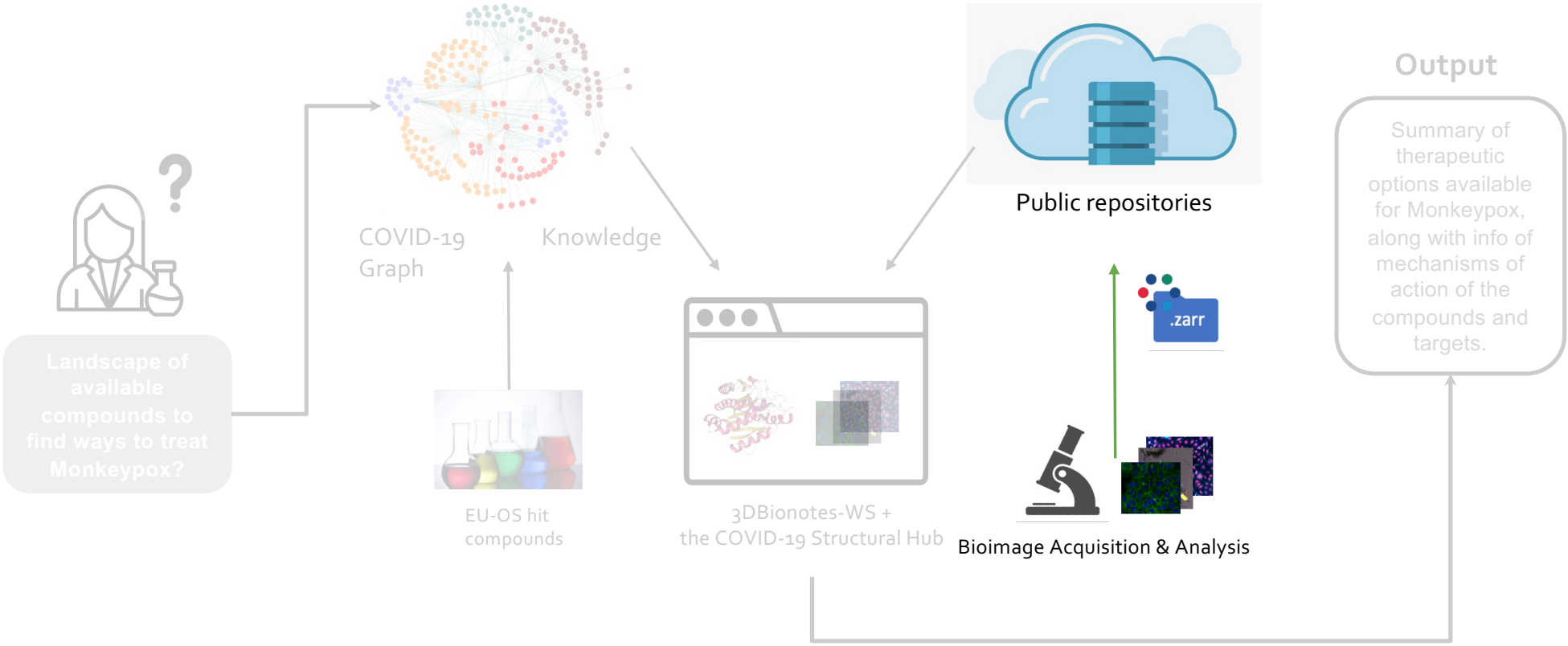
The demonstration – 3DBionotes-WS workflow



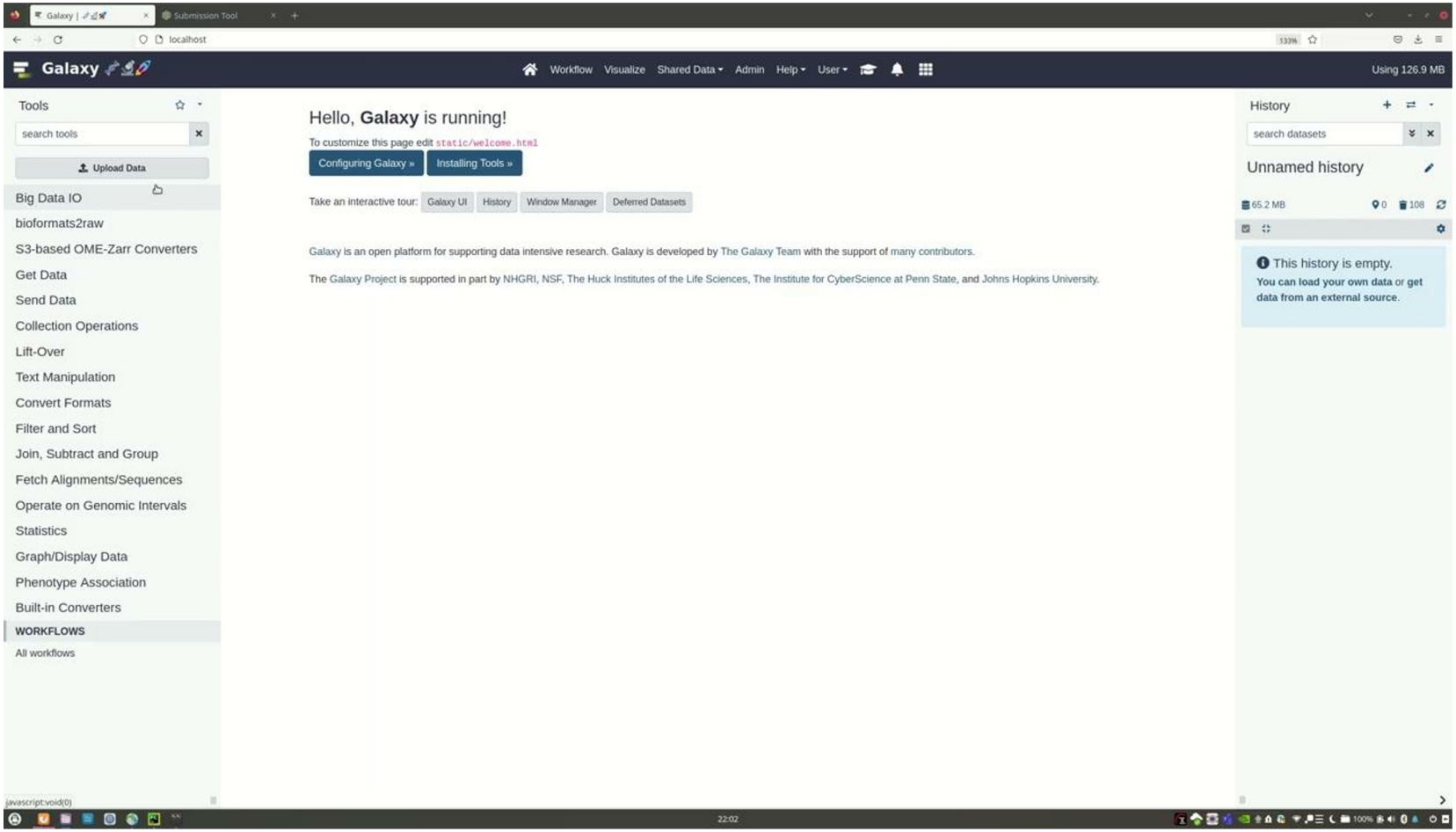
Dissemination of the 3DBionotes-WS

- Journal titled "**3DBionotes COVID-19 edition**"
  - https://doi.org/10.1093/bioinformatics/btab397
- Onboarded in EOSC marketplace
  - https://marketplace.eosc-portal.eu/services/3dbionotes-ws/
- Poster presentation at ELIXIR and EOSC symposium 2022
- Webservice
  - https://3dbionotes.cnb.csic.es/ws

The demonstration – OME-Zarr workflow



The demonstration – OME-Zarr workflow



Galaxy tools developed as part of the science project:

Big Data IO: A set of wrappers based on open-source FTP and Aspera clients to facilitate data transfer between Galaxy history and Bioimage Archive

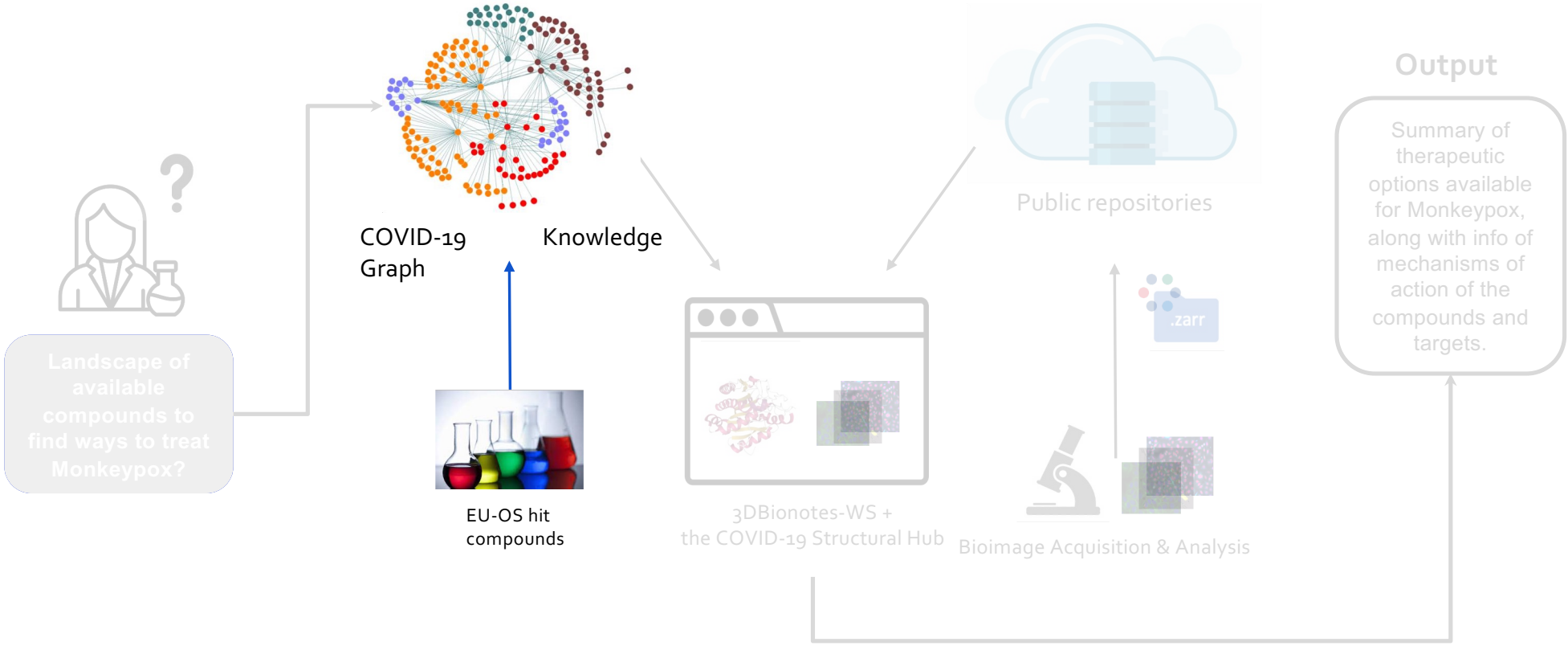
bioformats2raw: Tool based on the NGFF conversion package bioformats2raw plus metadata parsers to facilitate HCS data inputs

S3-based OME-Zarr Converters: Tool to convert datasets directly in the s3 object store without explicit data transfer

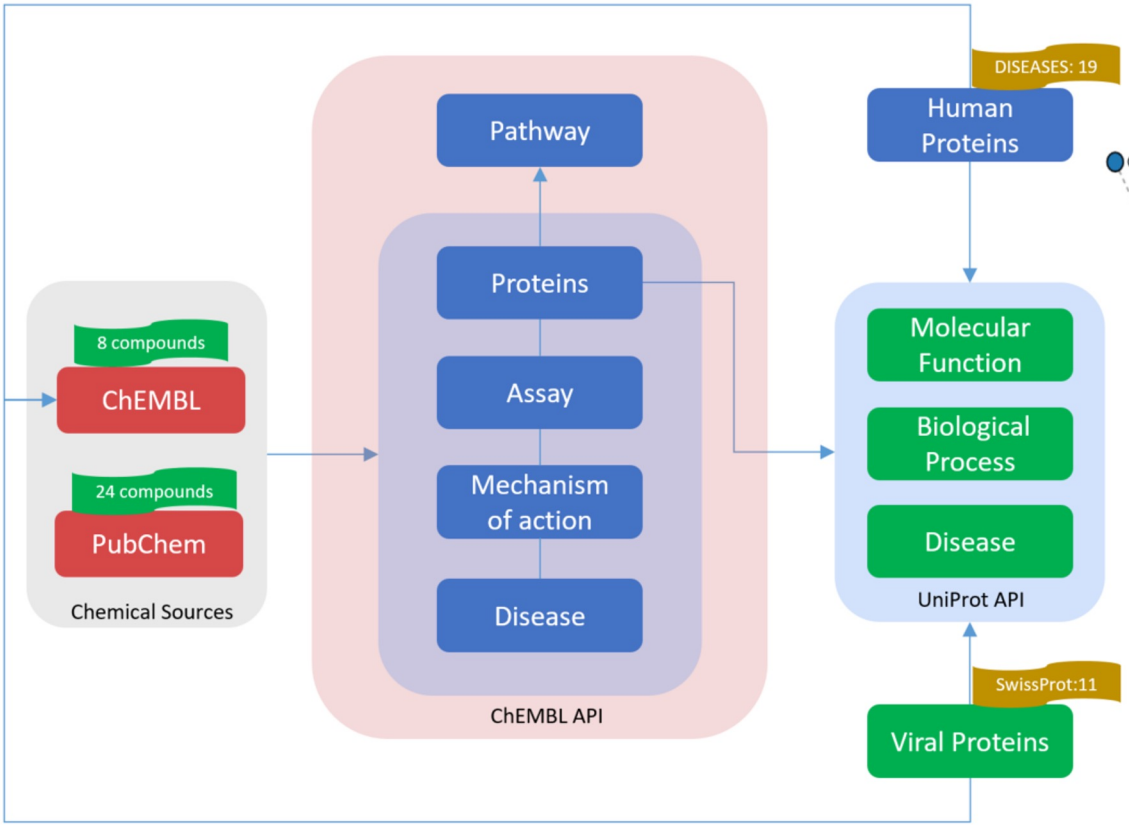
Dissemination of the OME-Zarr Workflow

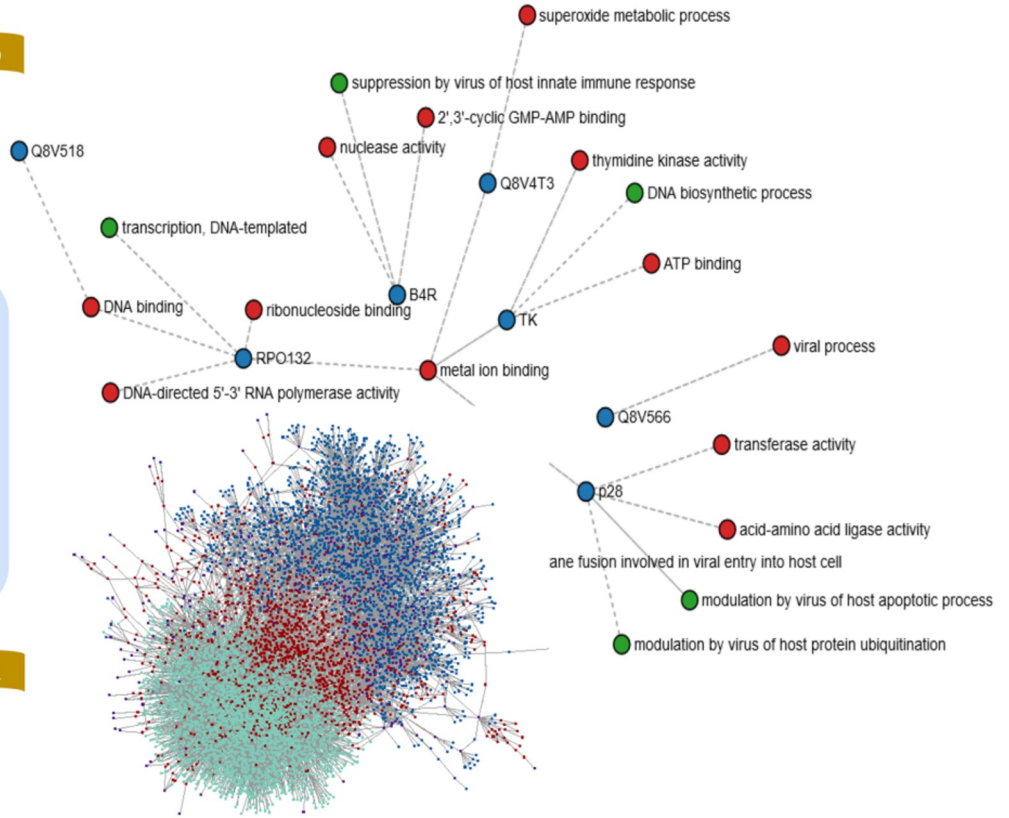
- Publishing the developed Galaxy tools to ToolShed and the workflow to WorkflowHub
- Giving courses on OME-Zarr data format and the relevant tooling:
  - NEUBIAS EOSC-Life Bioimage Analysis in the Cloud Workshop
- Participation in hackathons and workshops:
  - https://globalbioimaging.org/international-training-courses/bioimage-hackathon-01
  - https://galaxyproject.org/events/2022-10-egd

The demonstration – KG workflow

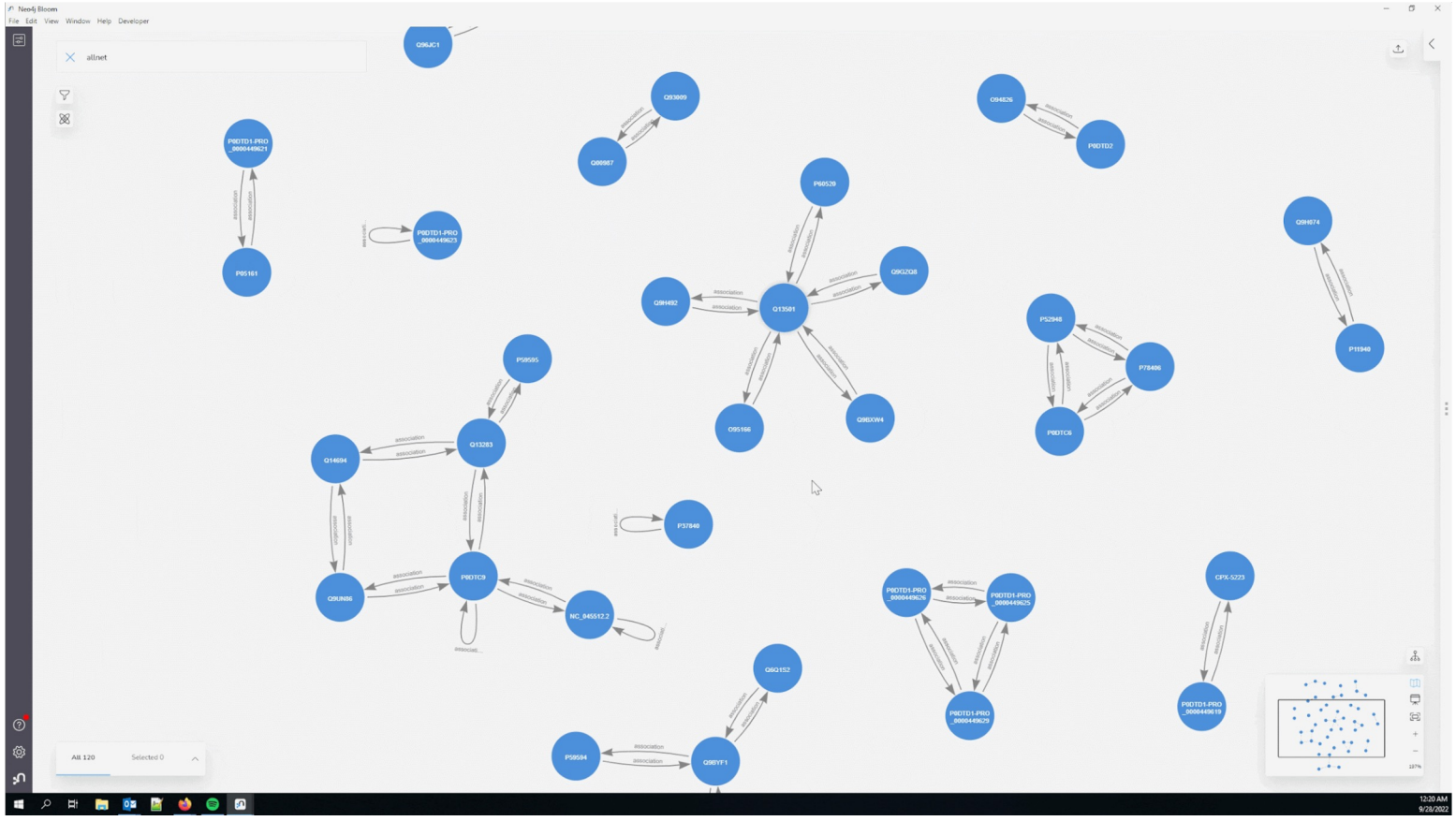


The demonstration – KG workflow





Proteins in KG annotated with 3DBionotes



Proteins in KG annotated with 3DBionotes

Proteins in KG annotated with 3DBionotes

KG dissemination

- GitHub: https://github.com/Fraunhofer-ITMP/mpox-kg
- Manuscript submitted to Bioinf. Advances:
  - https://www.biorxiv.org/content/10.1101/2022.08.02.502453v1
- Other KG applications in a journal paper
  - Title: **Comprehensive Fragment Screening of the SARS-CoV-2 Proteome Explores Novel Chemical Space for Drug Development**
    - https://doi.org/10.1002/ange.202205858
- BioModels: https://www.ebi.ac.uk/biomodels/MODEL2208040001
- Poster presentation at EOSC symposium 2022
- Available at COVID-19 Data Portal

Impact of KG

- General workflow for infectious disease/possible threats in place
- KG workflow was initially created for COVID-19
  - Monkeypox KG and Marburg KG were created in a week!
- Received interest from a national health security agency for using KG
  - Assisted them in deploying KG
- Compliant with FAIR annotations
  - Allows seamless transformation/integration to/with other formats
    - Collaboration with University of Luxembourg in BY-COVID project

Q & A – Demonstration:

Scientific demonstration

The EOSC Future project is co-funded by the
European Union Horizon Programme call
INFRAEOSC-03-2020, Grant Agreement 101017536