

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**, **highdimensional 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.

EOSC Future



The Scientific case in a nutshell: TSP Action



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).

EOSC Future

• Centralize and unify data exploiting and sharing among the different communities (EOSC Horizontal Services).

@EOSCFuture

EOSCfuture



eoscfuture.eu

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:

eoscfuture.eu

 Galaxy workflow for image data conversion to OME-Zarr and submission to Bioimage Archive is to be deployed to WorkflowHub

@EOSCFuture in EOSCfuture

EOSC Future

• COVID-19 KG (<u>https://github.com/Fraunhofer-</u> ITMP/BY-COVID-KG)

- mpact
- 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.









In production





Cell L	nes siRNAs Antibodies Compounds Organisms About	Search:
IN		General Accusition Preview 96
	index: Field#1 v	00
Explore lags Shares		Full viewer
		H5
idr0094-ellinger-sarscov2	- , :괜찮!! (2017); [- 2:: : : : : : : : : : : : : : : : : :	
ESP0025769	수 없는 것 없는 것이 집에 가 없다. 젖 집에는 바람이 없다. 것 같아. 않는 것이 없는 것이 없는 것이 없다.	Well ID: 1557054 Owner: Public data Show all V
ESP0025770		
ESP0025771		Attributes 6
ESP0025773	- 11월 - 11일 - 11	Organiem + P 🕯 🗙
ESP0025774		Added by: Public data
ESP0025776	한 것, 이 것과 같았다. 것 같아. 이 것 이 비슷하지 않아야 한 것, 이 야 같을 살	Organism Homo sapiens
B ESP0025777	- 2 1월 - 1월 - 1일 - 1일 - 1일 - 1일 - 1일 - 1일 -	
ESP0025778		Others
ESP0025779		Added by: Public data Quality Control pass
ESP0025841	요즘 것 같아요. 한 것 같아요. 이 지지 않는 것 같아요. 이 집에 가지 않는 것 같아요. 이 것 않아요. 이 집 않아요. 이 것 않아요. 이 것 않아요. 이 집 않	Percentage Inhibition (DPC) 9,55
ESP0025842		Hit Compound (over 75% no
ESP0025843		Number of Cells (DPC) 8867
ESP0025844		Phenotype Annotation Level well level
ESP0025845		Channels Cell body
GUE0000001		
GUF0000002		Tables 1
GUF0000003		Attachments 0
GUF0000004		
GUF0000005	그는 사장에 가장하는 사장은 것이다. 말한 책 동안에 앉아서, 동안에 앉았다. 소문을 통했다.	Comments 0
GUF0000006		Tage 0
GUF0000007		Tags 0
GUE000008	수 것 같아요. 한 것 않는 것 같아. 한 것 않는 것 같아. 한 것 않는 것 같아. 것 같아. 것 같아.	Ratings 0
Field positions in well x		Others 0
F	Zoom:	
	wrap () images per row Fields from well	
	H5	
	Zoom:	

Dissemination of the 3DBionotes-WS

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



The demonstration – OME-Zarr workflow



The demonstration – OME-Zarr workflow

		3277) 14	0 2 -
📕 Galaxy 🖉 💆	🎢 Workflow Visualize Shared Data - Admin Help - User - 🖝 🌲 🏢		Using 126.9 ME
Tools 🌣 *	Hello Galayy is running	History	+ = -
search tools	To customize this page edit static/weicome.html	search datasets	* ×
1. Upload Data	Configuring Galaxy » Installing Tools »	Unnamed histo	ry /
ig Data IO	Take an interactive four Galaxy III History Workey Manager Datasets	Printer	
oformats2raw	Toko or menuturo toon. Oleady or Talady History manager Deteried valuates	■ 05.2 MB	V0 108 2
based OME-Zarr Converters	Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.		
et Data	The Galaxy Project is supported in part by NHGRI. NSF. The Huck Institutes of the Life Sciences. The Institute for CyberScience at Penn State, and Johns Hookins University.	This history is empty.	
end Data		data from an extern	al source.
ollection Operations			
ft-Over			
ext Manipulation			
onvert Formats			
Iter and Sort			
in, Subtract and Group			
tch Alignments/Sequences			
perate on Genomic Intervals			
atistics			
aph/Display Data			
nenotype Association			
ilt-in Converters			
ORKFLOWS			
workflows			
crist-sold/0			,

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

EOSC Future



eoscfuture.eu

Dissemination of the OME-Zarr Workflow

- Publishing the developed Galaxy tools to <u>ToolShed</u> and the workflow to <u>WorkflowHub</u>
- 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:
 - <u>https://globalbioimaging.org/international-training-courses/bioimage-hackathon-o1</u>
 - o <u>https://galaxyproject.org/events/2022-10-egd</u>









The demonstration – KG workflow



Proteins in KG annotated with 3DBionotes















- GitHub: <u>https://github.com/Fraunhofer-ITMP/mpox-kg</u>
- Manuscript submitted to Bioinf. Advances:
 - <u>https://www.biorxiv.org/content/10.1101/2022.08.02.502453v1</u>
- 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





- 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