

## cryo Electron Microscopy in the European Open Science Cloud

*Tuesday, March 21, 2023 3:00 PM (30 minutes)*

In these pandemic times our group has coordinated large National and International Consortia to understand, through cryo Electron Microscopy (cryo-EM), both key issues on SARS Cov 2 spike dynamic (Melero et al., IU-CrJ. 2020) as well as specific properties of mutations that were prevalent in Europe at certain periods (Ginex et al., PLoS Pathog. 2022), as part of our work in the European Research Infrastructure for Integrative Structural Biology, Instruct-ERIC. These works have been complemented with others in Bioinformatics as developers of one of the few ELIXIR Recommended Interoperability Resources (Macias et al., Bioinformatics. 2021). Along this hard work we have learnt many lessons and realized many new needs that are now guiding our efforts in cloud computing and in data management. In the following we will briefly review some of these new developments.

The first one is termed “ScipionCloud”, and it is a service registered in the EOSC Marketplace where users from the Instruct Research Infrastructure can deploy a cluster in the cloud to process the data acquired at an Electron Microscopy facility. This cluster has all cryoEM packages and software needed to obtain a 3D structure and is powered by EOSC agreed-on computing resources on the back-end. This means that scientists with minimal computational background (or compute resources of their own) can access the latest tools as well as powerful computational resources to obtain a refined 3D structure to be published and shared with the community. The service quality is assessed through the SQAaaS utility, developed in the same project, that allows to check different quality metrics in software development projects. In addition, the tool permits to evaluate the FAIRness of service outputs that are stored in public repositories. Finally, minimal modifications of the service are needed to deploy a similar cluster in the AWS cloud. Documentation of this service can be found in the Scipion website (<https://scipion.i2pc.es/>) and in arXiv:2211.07738

The second development addresses the issue of lack of standardization, specially in the area of information (Image) processing, trying to improve the FAIRness of cryoEM workflows. In this way we have developed the tools to export the image processing workflow in Common Workflow Language, using a CryoEM ontology and depositing workflows in WorkflowHub. On this front we have published a cryoEM ontology in the following catalogues: Ontology Lookup Service (<https://www.ebi.ac.uk/ols/ontologies/cryoem>), BioPortal (<https://bioportal.bioontology.org/ontologies/CRYOEM>) and FAIRsharing (10.25504/FAIRsharing.q47I0t). Additionally, we have developed a Scipion plugin to publish processing workflow template in WorkflowHub in the form of a RO-Crate object with Scipion JSON and CWL workflow (enriched with previous ontology) + diagram + metadata.

On a further effort to first map/understand the current situation on raw data deposition in public databases, we first performed an analysis of the pre pandemic situation and the current one in terms of deposition of cryoEM data, and we were surprised to see that, essentially, nothing had been learnt through the pandemic, and that there was a very widespread lack of raw data structural data, with also substantial differences among “Regions” (Europe deposits around 10% of the acquired data, the US 2%, and virtually none from Asia). To act on this situation, an ambitious push towards data sharing is being performed by Instruct-ERIC, coordinated from this laboratory. Still at the pilot stage, we are developing a strategy across the entire infrastructure, starting at the individual facilities which are to archive quality annotated data and share the data in a Federated manner using either Onedata or IRODS (both solutions are under testing). The whole interplay among Facilities will be orchestrated from Instruct Hub through new extensions of its project management system ARIA. In short, almost three intense years of work in the context of the current SARS Cov2 pandemic, marked by the need to increase efficiency and be better prepared for the likely events of future pandemics.

**Primary authors:** Prof. SORZANO SANCHEZ, Carlos Oscar (CNB-CSIC, Instruct-ERIC); Ms SANCHEZ, Irene (CNB-CSIC, Instruct-ERIC); Ms DEL CANO, Laura (CNB-CSIC, Instruct-ERIC); CARAZO GARCIA, Jose Maria (National Center for Biotechnology - CNB - CSIC)

**Presenter:** CARAZO GARCIA, Jose Maria (National Center for Biotechnology - CNB - CSIC)

**Session Classification:** Health & Life Sciences (including Pandemic Preparedness Applications)

**Track Classification:** Track 2: Health & Life Sciences (including Pandemic Preparedness Applications)