A FAIR and Cloud-enabled approach
for Electron Microscopy data processing

Laura del Cano[1], Irene Sánchez[1], José María Carazo[1], Carlos Oscar Sorzano[1], Tomas Svoboda[2], Radek Veverka[2], Josef Handl[2], Ales Krnek[2], Jirka Novacek[2]

[1] National Center of Biotechnology (CNB – CSIC)
[2] Central European Institute of Technology – Masaryk University (CEITEC – MU)

Thanks to the integration with EOSC resources, CryoEM data can now be processed in cloud infrastructures while ensuring that data is kept properly annotated and eventually be made public in compliance with the FAIR principles promoted in Open Science.

FAIR data from the microscope

1 An Instruct-ERIC user is granted a project on a microscope facility. The project details are in ARIA (Instruct-ERIC Access Management System).
In the facility, details of the sample and session are stored in a LIMS (Laboratory Information Management System) server.

2 The facility operator starts the acquisition session and while data is acquired it is processed using Scipion in streaming mode. The processing workflow can be included in the metadata.

3 Both raw data and collected metadata are sent to a Onedata provider located at the facility and registered in the EGI Datahub.
Data location is added to the LIMS entry and the ARIA proposal but they will be kept private until the Instruct embargo period is over.
The user can start downloading the raw data while the session is still ongoing or wait until it is completed.

How Onedata data management platform works?

Data is organized into distributed virtual volumes (**spaces**), which can be supported by several entities providing the physical storage (**providers**) federated in a particular **zone**
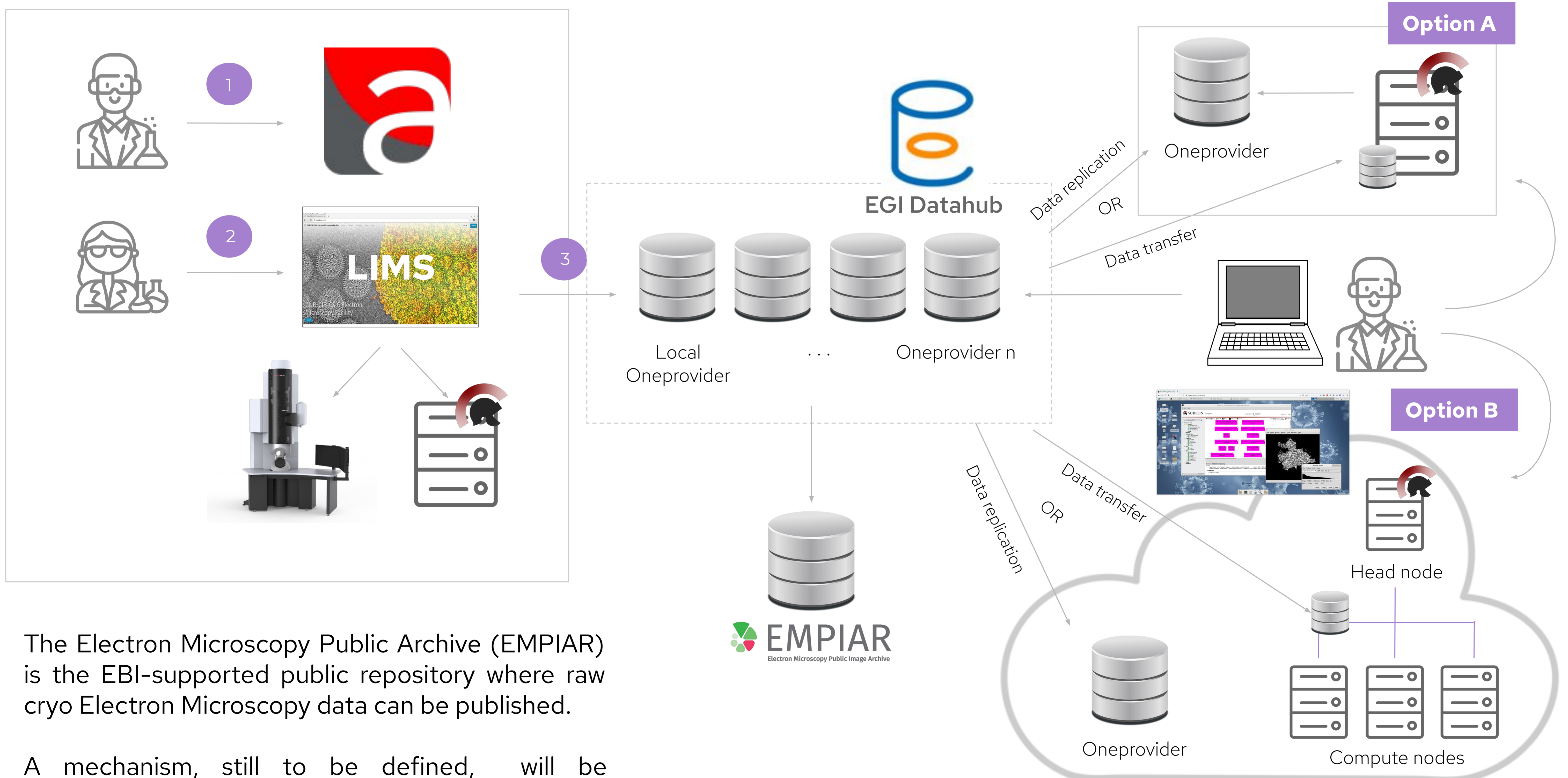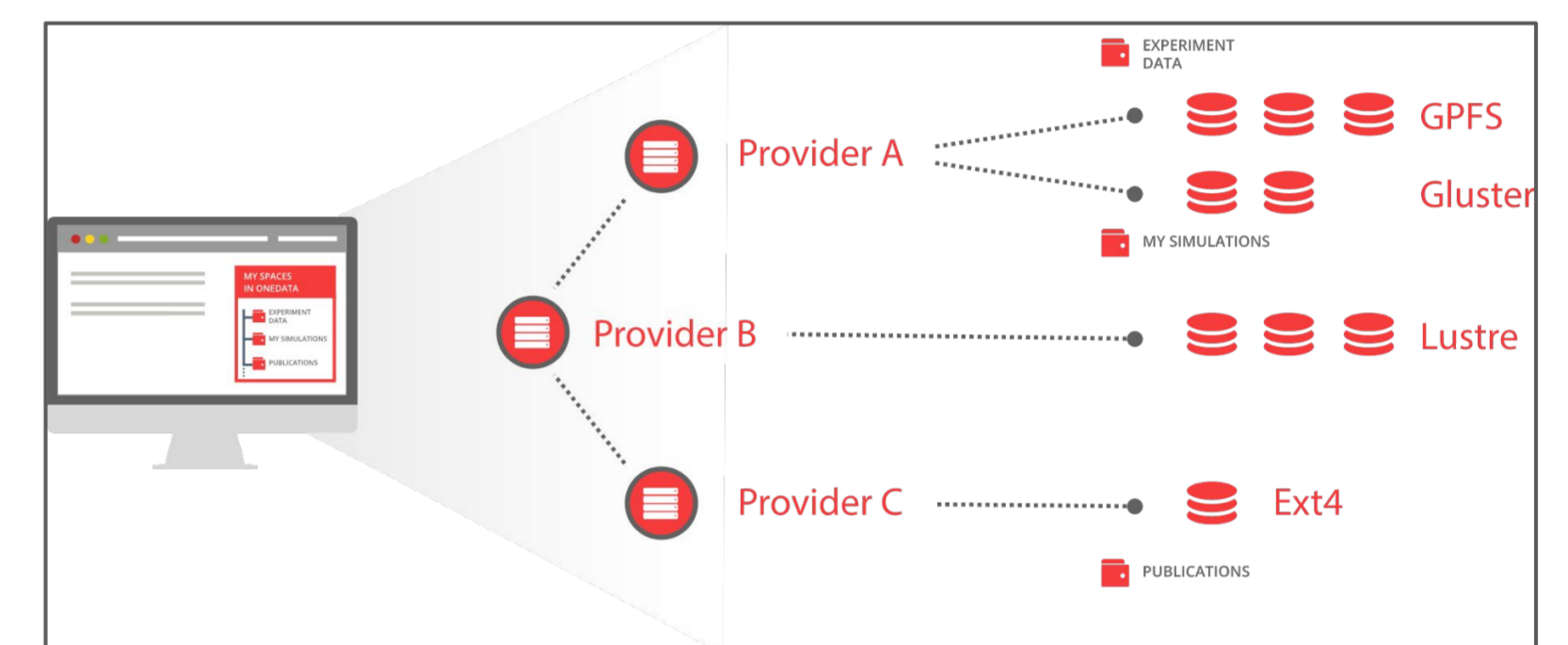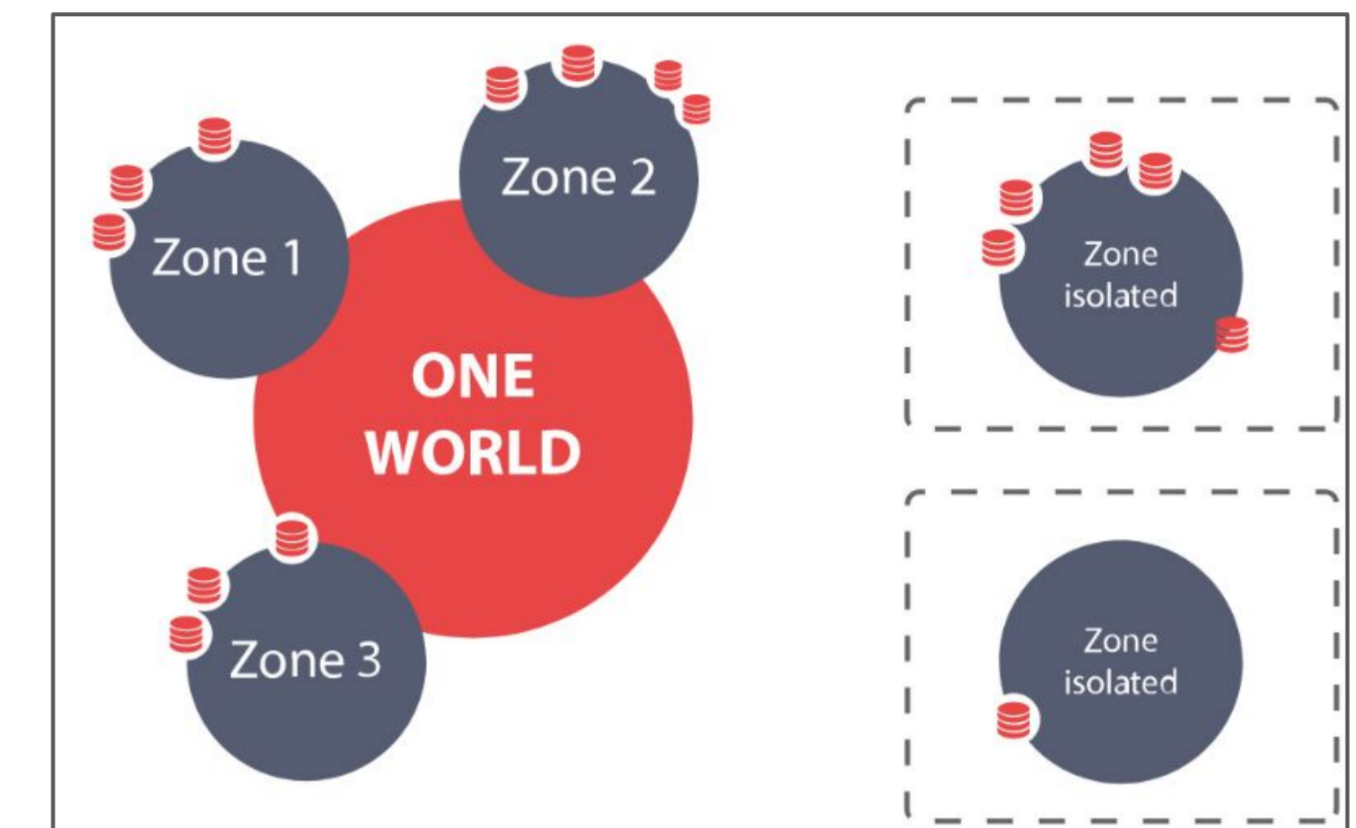
The Electron Microscopy Public Archive (EMPIAR) is the EBI-supported public repository where raw cryo Electron Microscopy data can be published.

A mechanism, still to be defined, will be implemented to publish datasets from the OneData space to EMPIAR where it will be openly available.

Data processing in the cloud

The user can download the acquired data to process it locally (A) or request to use the ScipionCloud service in the EOSC Portal that will deploy a virtual cluster on EOSC Compute resources or public clouds such as AWS (B).
In both cases data can be either transferred to the local disk or replicated to a Oneprovider in the same network and mounted in the server or cluster.