Workflows co-working sprint finish

Nicolas Soler - head of scientific Data Management

Alba synchrotron

Why workflow engines?

- reproducibility
- scalability: easy parallelization (from laptop to HPC cluster)
- containerization
- control of intermediate processes

→ **put your scripts in processes and chain them together**

nextflow https://www.nextflow.io/

Where do we want to use them?

Tomography

- μ-tomography (future FAXTOR beamline)
- cryo-nanotomography (MISTRAL beamline)

crystallography

- MX (XALOC beamline)
- SSX (future XAIRA beamline)

FaXToR – Mistral status

Tomo stack deconvolution workflow.

This workflow has **7 steps**:

- File conversion to hdf5 » Normalization » Angle extraction »
- File conversion to mrc » Wiener Deconvolution »
- Absorbance calculation » File conversion to hdf5

This workflow need to mix bash scripts with python scripts.

We are able to chain some processes and pass parameters to the workflow.

Work in progress

MX workflow

Characterization workflow

- capture a few diffraction images
- locate spots (Bragg peaks) and their quality
- evaluate the point group
- propose a data collection strategy

In summary

This coworking sprint was the occasion to:

- use Nextflow outside of its computational biology context
- have our data engineers getting familiar with it
- establish a close and durable contact with Nextflow mentors
- start redesigning and modernizing our pipelines