



Contribution ID: 56

Type: **Oral presentation**

Jupyter for Reproducible Science at Photon and Neutron Facilities

Tuesday, 15 October 2019 10:10 (20 minutes)

Modern photon and neutron facilities produce huge amounts of data which can lead to interesting and important scientific results. However the increasing volume of data produced at these facilities leads to some fundamental issues with data analysis.

With data sets in the hundreds of terrabytes it is difficult for scientists to work with their data, the large size also leads to another issue as these huge volumes of data require a lot of computational power to be analysed, lastly it can be difficult to find out what analysis was performed to arrive at a certain result, making reproducibility challenging.

Jupyter notebooks potentially offer an elegant solution to these problems, they can be ran remotely so the data can stay at the facility it was gathered at, and the integrated text and plotting functionality allows scientists to explain and record the steps they are taking to analyse their data, meaning that others can easily follow along and reproduce their results.

The PaNOSC (Photon and Neutron Open Science Cloud) project aims to promote remote analysis via Jupyter notebooks, with a focus on reproducibility and following FAIR (Findable, Accessible, Interoperable, Re-Usable) data standards.

There are many technical challenges which must be addressed before such an approach is possible, such as recreating the computational environments, recording workflows used by the scientists, seamlessly moving these environments and workflows to the data, and having this work through one common portal which links together several international facilities.

Some of these challenges have likely also been encountered by other scientific, as such it would be very useful to work together and try to come up with common workflows and solutions to creating reproducible notebooks for science.

This project (PaNOSC) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852.

Primary author: ROSCA, Robert (European XFEL)

Presenter: ROSCA, Robert (European XFEL)

Track Classification: Workshop: Building reproducible workflows for earth sciences