Contribution ID: **89** Type: **Presentation**

# Driving data analysis through the Jupyter Notebook at European XFEL

*Wednesday, 10 October 2018 17:30 (15 minutes)*

Computational science based on simulation or experimental data typically requires data analysis to extract insight from potentially large data sets. In this project, we explore the suitability of the Jupyter Notebook to drive the processing chain from raw data to figures used in publications and reports.

Computational Science is emerging as a key tool in academia and industry. For example, in the field of magnetism, simulations of nano structures have become well established and are used widely. In photon science, the analysis of experimental data is essential and central to understanding correlations, adjusting experiment parameters, and exploiting an instrument's full potential.

With regards to this increasing importance in science, there is increasing concern about the reproducibility of scientific results obtained mainly from computational data analysis [1]. Ideally, any scientist should be able to recreate, for example, central figures in publications. This requires keeping track and publishing of all steps taken during the analysis, including tracking of all experiment and simulation runs, data and simulation results used from each, and all metadata, parameters and processing steps.

We study the utility of the Jupyter Notebook as a virtual research environment for this common scenario through both data analysis in computational modelling of magnetic devices and Photon Science. For the former, we have taken a well established micromagnetic simulation package [2] based on C++ and added a Python interface [3] to allow convenient control of the package through the Jupyter Notebook [4].

Of particular interest for both application domains is that within the Jupyter Notebook, we can carry out simulation, data analysis, and specialised post-processing within a single document, making the work more easily reproducible and distributable. A special case is the creation of figures in publications: by creating each central figure in a publication within a Jupyter Notebook, we can publish the notebooks together with the manuscript, and thus make the key data elements of the publication reproducible.

Emerging developments such as the European Science Cloud (EOSC) demand that the whole computational analysis process can be driven remotely. The method of driving computational science through the Jupyter Notebook provides the remote execution elegantly: by hosting the Jupyter Notebook server where the data and simulation capability is, and connecting the user's web browser with the Jupyter Notebook server via HTTPS, we avoid common problems experienced with remote desktops or X forwarding. Driving computational analysis through the Jupyter Notebook can provide a flexible cloud-enabled data analysis infrastructure.

This project is part of the Jupyter-OOMMF activity in the OpenDreamKit [5] project and we acknowledge the financial support from Horizon 2020 European Research Infrastructures project (676541). The work is also supported by the EPSRC CDT in Next Generation Computational Modelling EP/L015382/1.

[1] M. Baker, Nature 533, 452 (2016).
[2] M. J. Donahue and D. G. Porter, OOMMF User's Guide, Version 1.0, Interag. Rep. NISTIR 6376, NIST Gaithersburg (1999).
[3] M. Beg et al. AIP Advances 7, 056025 (2017).
[4] https://github.com/joommf
[5] https://opendreamkit.org

## Type of abstract

Presentation

## Summary

We present the use of the Jupyter Notebook as a virtual research environment. We focus on reproducible data analysis in computational science in the areas of Photon Science and micromagnetic modelling, and highlight opportunities for use of this e-infrastructure in the context of the European Open Science Cloud.

**Primary author:** Dr BEG, Marijan (European XFEL GmbH)

**Co-authors:** PADEE, Anna (European XFEL GmbH); Prof. FANGOHR, Hans (European XFEL GmbH); Mr PEPPER, Ryan (University of Southampton); Dr BROCKHAUSER, Sandor (European XFEL GmbH); HAUF, Steffen (European XFEL GmbH); Dr KLUYVER, Thomas (European XFEL GmbH); JAROSIEWICZ, Tobiasz (European XFEL GmbH)

**Presenter:** Dr BEG, Marijan (European XFEL GmbH)

**Session Classification:** Computing Services Part II

**Track Classification:** Area 3. Computing and Virtual Research Environments