

Proposal to run a BoF on Notebooks at RDA Helsinki

TL;DR

A proposal is being put together to run a BoF session on Notebooks (i.e. discussions on Jupyter and myBinder but if this is too narrow a scope please say so) at RDA Helsinki which will run from 23rd-25th of October. The deadline for the proposal is the 27th June.

If you are interested in getting involved, attending the meeting or have suggestions for discussion topics, then please either email

Hugh dot Shanahan at rhul dot ac dot uk or do a pull request on this document.

Introduction

Notebooks, specifically Jupyter notebooks (but also notebooks based on Rmarkdown) are generating a great deal of excitement and are potentially a significant step forward in terms of reproducibility, education, code documentation and academic publishing. A recent(Pimentel et al., n.d.) paper found nearly 1.2 Million unique notebooks just going through github.

More recently the deployment of sites such as JupyterHub, myBinder and EGI notebooks mean that notebooks can be deployed on a cloud and hence obviates issues with installed libraries. One potential result of this is that it would represent the easiest pathway for users to use cloud computing resources for research which could transform the use of such resources. It is important that initiatives such as EOSC become more aware of such initiatives at a policy level.

Proposal

Details of how to apply for a BoF can found here. The application is itself relatively short though it's worth looking carefully at the criteria that are set. It's important to make sure it matches the RDA's mission and that it's not simply a set of talks but that there's room for discussion.

Possible discussion points for BoF

The following are a set of possible topics that could be discussed at the BoF. There probably is time to discuss

Notebook citation

Given the profusion of notebooks it is clear that a system of enabling the citation of notebooks needs to be implemented. In many respects this represents an opportunity as the infrastructure behind notebooks is comparatively new and hence potentially easily adapted to follow citations guidelines for the Force 11 software citation implementation group and the software source code Interest Group citation group in the RDA.

Influencing policy

The EGI already successfully deployed a notebook solution. Can the functionality of Binder be similarly deployed? How do notebooks fit into the EOSC vision and other similar platforms? If not what steps can be made to influence thinking in this area.

Software engineering and notebooks

A valid criticism of notebooks are that it encourages poor software engineering practice such as TDD or encapsulating the notebooks into libraries. Some of this stems from the fact that notebook editors are nowhere near as developed as other IDE's but the first concerns are genuine. What steps can be taken to improve this situation?

Pimentel, Joao Felipe, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. n.d. "A Large-Scale Study About Quality and Reproducibility of Jupyter Notebooks," 11.