Contribution ID: **22**   Type: **Presentation**

JupyterLab+ScienceMesh: Collaborative Data Science in sync-and-share environment.

Wednesday, 26 January 2022 11:20 (20 minutes)

Collaborative Data Science becomes increasingly important, as organizations continue to become more data-driven, and Data Science projects/models become more complex. In the report **Critical Capabilities for Data Science and Machine Learning Platforms** (March 2021) Gartner predicts, that in near future collective intelligence in Data Science and cloud-based AI infrastructure will be among key factors for competitive advantage.

This talk presents Distributed Data Science environments (part of ScienceMesh), which allow collaboration on Jupyter Notebooks in sync-and-share environment.

Jupyter Notebook has become No1 platform used by data scientists to build interactive applications and to work with big data and AI. It is widely used in CS3 institutions, many successful applications have been presented in CS3 conferences.

ScienceMesh, developed in CS3MESH4EOSC project, creates the Federated Scientific Mesh providing federated sharing of data across different sync-and-share services, federated use of applications (such as collaborative document editing, data archiving, and data publishing), fast transfer of large datasets and remote data analysis (Data Science environments).

For Data Science environments ScienceMesh delivers a JupyterLab extension, integrating JupyterLab environment with ScienceMesh. File browsing and additional share and collaboration functionalities for notebooks and resources across federated cloud are now possible in JupyterLab environment. JupyterLab is considered a complete, full-fledged IDE for Data Science tasks and interactive computing, where data scientists can do all their work in one tool, so the point is that functionalities for sharing (full cs3apis client) and concurrent editing are available inside this environment. On the other hand, Data Science environments are integrated with a comprehensive suite of Data Services in ScienceMesh, to support complete research and Data Science workflows with the use of existing collaboration tools.

The relevance and benefits of ScienceMesh Data Science Environments will be discussed in the context of two scientific use cases (High Energy Physics and Earth Observation), along with various business-related scenarios.

Primary author: SIEPRAWSKI, Marcin (Software Mind)

Presenter: SIEPRAWSKI, Marcin (Software Mind)

Session Classification: User Stories

Track Classification: Main session: User Voice: Novel Applications, Data Science Environments & Open Data