



Contribution ID: 210

Type: **Presentation**

## Data transfer, synchronization and sharing solution for PSDI

*Friday 21 March 2025 12:45 (15 minutes)*

PSDI is the UK nationally funded programme that analyses physical sciences needs in a common data infrastructure and develops guidance, training and technology to address these needs. The PSDI main objective is to serve research use cases originating in experimental “bench science” and simulations with applications in physics, chemistry, materials research or engineering. The main challenge to address by PSDI in the well-known “three Vs” of Big Data: Volume, Variety and Velocity is not Volume but primarily Variety aspect with some considerations given to Velocity, too.

Data transfer, synchronization and sharing solution is a part of broader technology works in PSDI and relies on Open Source components with a strong inclination to containerised and cloud deployments. The IT stack for developing the solution includes OCIS (Own Cloud Infinite Scale) with Ceph object store as a backend, combined with additional tools and an orchestration component based on Apache Airflow. We are reporting on integration of components, on performance measurements and on implementation of data policies that are essential to have in a common data infrastructure. We are discussing the potential of combining the data transfer, synchronisation and sharing solution with data pipelines and a data indexing solution that are also in scope of PSDI technology works.

[1] PSDI –Physical Sciences Data Infrastructure. [www.psdi.ac.uk](http://www.psdi.ac.uk)

[2] Towards data sharing service for Physical Sciences Data Infrastructure. CS3 2024. <https://indico.cern.ch/event/1332413/contributions/5>

[3] Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management. Volume 35, Issue 2, April 2015, Pages 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

**Author:** BUNAKOV, Vasily (STFC UKRI)

**Co-authors:** BELOZEROV, Alexander (STFC UKRI); PAWULA HEWAGE, Amali (STFC UKRI); WRIGHT, Paul (STFC UKRI); MEACHAM, Stuart (STFC UKRI); UNDERWOOD, Tom (STFC UKRI)

**Presenter:** PAWULA HEWAGE, Amali (STFC UKRI)

**Session Classification:** Data sharing infrastructures

**Track Classification:** Main sessions: CS3 federations and synergies with eResearch infrastructures.