

# Unified data access to e-Infrastructure, Cloud and personal storage within INDIGO-DataCloud

*Tuesday, October 11, 2016 11:30 AM (15 minutes)*

Nowadays users have a variety of options to get access to storage space, including private resources, commercial Cloud storage services as well as storage provided by e-Infrastructures. Unfortunately, all these services provide completely different interfaces for data management (REST, CDMI, command line) and different protocols for data transfer (FTP, GridFTP, HTTP). The goal of the INDIGO-DataCloud project is to give users a unified interface for managing and accessing storage resources provided by different storage providers and to enable them to treat all that space as a single virtual file system with standard interfaces for accessing and transfer, including CDMI and POSIX. This solution enables users to access and manage their data crossing the typical boundaries of federations, created by incompatible technologies and security domains. INDIGO provides ways for storage providers to create and connect trust domains, and allows users to access data across federations, independently of the actual underlying low-level storage technology or security mechanism. The basis of this solution is the Onedata platform (<http://www.onedata.org>). Onedata is a globally distributed virtual file system, built around the concept of "Spaces". Each space can be seen as a virtual folder with an arbitrary directory tree structure. The actual storage space can be distributed among several storage providers around the world. Each provider gives the user support for each space in a fixed amount and the actual capacity of the space is the sum of all declared provisions. Each space can be accessed and managed through a web user interface (Dropbox-like), REST and CDMI interfaces, command line as well as mounted directly through POSIX. This gives users several options, the major of which is the ability to access large data sets on remote machines (e.g. worker nodes or Docker containers in the Cloud) without pre-staging and thus interface with existing filesystems. Moreover, Onedata allows for automatic replication and caching of data across different sites and allows cross-interface access (e.g. S3 via POSIX). Performance results covering selected scenarios will also be presented.

Besides Onedata, as a complete monolithic middleware, INDIGO offers a data management toolbox allowing communities to provide their own data handling policy engines and delegating the actual work to dedicated services. The INDIGO portfolio ranges from multi tier storage systems with automated media transition based on access profiles and user policies, like StoRM and dCache, via a reliable and highly scalable file transfer service (FTS), with adaptive data rate management to DynaFed, a lightweight WebDAV storage federation network. FTS is in production for over a decade and is the workhorse of the Worldwide Large Hadron Collider Computing GRID. The DynaFed network federates WebDAV endpoints and lets them appear as a single overlay filesystem.

## Secondary Keyword (Optional)

Storage systems

## Primary Keyword (Mandatory)

Distributed data handling

## Tertiary Keyword (Optional)

**Primary authors:** KRYZA, Bartosz (ACC Cyfronet-AGH); DUTKA, Lukasz (Cyfronet); FUHRMANN, Patrick (DESY)

**Co-authors:** SALOMONI, Davide (INFN CNAF); DONVITO, Giacinto (INFN-Bari); DUTKA, Lukasz

**Presenter:** DUTKA, Lukasz (Cyfronet)

**Session Classification:** Track 4: Data Handling

**Track Classification:** Track 4: Data Handling