ONEDATA
Eventually Consistent Virtual Filesystem for Multi-Cloud Infrastructures

Michał Orzechowski (CYFRONET AGH)
PROBLEMS ADDRESSED BY ONEDATA

1. Multi-protocol transparent access to data “[…] but we want POSIX”
2. Heterogeneity of storage technologies
3. Replica Management
4. Easy Data Sharing and publication (DIO)
5. Metadata Management Integrated with Data Management Platform
6. Flexible authentication and authorization
7. Easy integration using API with external services
8. High-throughput data processing
PROBLEM 1: MULTI-PROTOCOL TRANSPARENT ACCESS TO DATA IN MULTI-CLOUD ENVIRONMENTS

- Transparently access and create data in multi-cloud environments
- Care less about data locality, all your data are accessible wherever you go
- Use many protocols to access the same data
ONEDATA FOR ONEWORLD

NO TRUST

Zone

Onezone

Entry GUI

REST APIs

OAI-PMH

Kademlia DHT (in prep.)

SAML (in prep.)

OIDC

SAML

NO TRUST

Zone

Oneprovider

Ceph

POSIX

PCIX

GridFTP (in prep.)

FTP / SFTP (in prep.)

POSIX

CDMI

S3

Swift

Data Mgmt. GUI

REST APIs

WebDAV (in prep.)

NO TRUST

NO TRUST
[...] BUT WE WANT POSIX

- Support for most of the POSIX operations on globally distributed virtual file system
- All data accessible via a unified file system mountable on virtual machines, Grid worker nodes and containers
PROTOCOL HANDLERS (PLUGINS)

Onezone

Entry GUI

Kademlia DHT (in prep.)

Oneprovider

POSIX

CDMI

Data Mgmt. GUI

REST APIs

FTP / SFTP (in prep.)

WebDAV (in prep.)

FUSE Client

Oneclient

HTTP GUI REST

FUSE Client

HTTP GUI REST

FUSE Client

HTTP GUI REST

FUSE Client
PROBLEM 2: HETEROGENEITY OF STORAGE TECHNOLOGIES

- Use the data protocols of your choice to access data wherever you go
- Minimize the problems of selection right storage technology to data centres operators
- Avoid cloud vendor locking
DIFFERENT TYPES OF STORAGES VIRTUALIZED

- Provider A
  - S3
  - POSIX
- Provider B
  - Ceph
- Provider C
  - OpenStack Swift
  - PUBLICATIONS

[Diagram showing connections between providers and storages]
STORAGE SYSTEMS DRIVERS (PLUGINS)

Onezone

Entry GUI

Kademlia DHT (in prep.)

Oneprovider

POSIX

Ceph

S3

Swift

GridFTP (in. Prep.)

FUSE Client

Oneclient

HTTP GUI REST

FUSE Client

HTTP GUI REST

FUSE Client

FUSE Client

ONE WORLD
PROBLEM 3: REPLICA MANAGEMENT

• Replicate files on demand and on the fly without any additional effort
• Migrate data between sites on demand with simple API interface
• Easily check location of your data using GUI or API
REPLICAS MANAGEMENT SIMPLIFIED

- Manage files not Replicas
- File distribution between storage locations is underneath the file structure
- Replicas management on a chunk basis
- Missing chunks delivered on the fly
- API for replica management for pre-staging and implementing external data policy management
PROBLEM 4: EASY DATA SHARING WITHOUT BORDERS

- Share large scale data collections with other communities
- Enable your data to be shared in cross-federation scenarios
- Bring your data and tools as building blocks to European Open Science Cloud
EASY DATA SHARING

- Team-sharing
  - For groups
  - For individuals
  - Token based
- Cross-community data sharing
- Instant and ad-hoc data sharing
- Thanks to effort supported by EGI Engage:
  - Open Data Publication
  - Handles (DOI) support
  - OAI-PMH
EASY DATA SHARING

- Team-sharing
  - For groups
  - For individuals
  - Token based
- Cross-community data sharing
- Instant and ad-hoc data sharing
- Thanks to effort supported by EGI Engage:
  - Open Data Publication
  - Handles (DOI) support
  - OAI-PMH
EASY DATA SHARING

- Team-sharing
  - For groups
  - For individuals
  - Token based
- Cross-community data sharing
- Instant and ad-hoc data sharing
- Thanks to effort supported by EGI Engage:
  - Open Data Publication
  - Handles (DOI) support
  - OAI-PMH
OxfordFlowerDatabase-FlowerSet1

Path: t7Y7mBQXgLVz2RCRG_nvF9dh268H86MvkJ-0Hm3j84 > FlowerSet1

Public URL: https://datahub.egi.eu/share/h7tiS2Dznk2RS8kCaA--PZG7y/

License: CC-O

Attribute: Value

Files:
- image_0001.jpg (50.85 KB, 2016-09-27 15:09)
- image_0002.jpg (41.24 KB, 2016-09-27 15:09)
- image_0003.jpg (46.2 KB, 2016-09-27 15:09)
- image_0004.jpg (30.16 KB, 2016-09-27 15:09)
A directory with path "\:Y7m8QXGlv2RCRG_nvF9dh268H86McKj-0Hm3j84/FlowerSet1" is shared with name "OxfordFlowerDatabase-FlowerSet1"

**Public URL**

You can give an access to the shared directory to anyone with the public URL:

https://datahub.egi.eu/share/h7tzi5D2NmK2R5BkCaoA-\_P2G7yARRXBFxFEnxjnWMs
**Open Data Platform Workflow**

1. **Register identifier (e.g. DOI)**
   - **DataCite**
   - Expose share metadata over OAI-PMH

2. **Create a share from folder or file with Metadata in DC**
   - **Share of Dataset**
   - Lazy Replication

3. **View data set preview in browser using DOI**
   - **Public Services For Data Discovery**
   - **OpenAIRE**
   - **Copy of Data-set-1.1**
   - **Copy data set**

4. **Discover data set**
   - **Private Resources**
   - **Dataset**

5. **View data set preview in browser using DOI**

6. **Private Resources**
OxfordFlowerDatabase-FlowerSet1

Path: t7Y7mBQXgLv2RCRG_nwF9dh268H86MicKj-0Hm3jB4 > FlowerSet1

Public URL: https://datahub.egi.eu/share/h7tziSD2NmK2RSbkCaoA__PZG7Y...

**BASIC**

**license**

CC-0

**Attribute**

**Value**

**Save all changes**  **Discard changes**  **Remove metadata**

---

**FlowerSet1**

**FILES**

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>image_0001.jpg</td>
<td>50.85 KB</td>
<td>2016-09-27 15:09</td>
</tr>
<tr>
<td>image_0002.jpg</td>
<td>41.24 KB</td>
<td>2016-09-27 15:09</td>
</tr>
<tr>
<td>image_0003.jpg</td>
<td>46.2 KB</td>
<td>2016-09-27 15:09</td>
</tr>
<tr>
<td>image_0004.jpg</td>
<td>30.16 KB</td>
<td>2016-09-27 15:09</td>
</tr>
</tbody>
</table>
Publish the share

Please enter Dublin Core metadata here:

```xml
<?xml version="1.0" encoding="UTF-8"?>

<metadata
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Oxford Flower Database</dc:title>
  <dc:creator>Maria Elena Nilsback</dc:creator>
  <dc:creator>Andrew Zisserman</dc:creator>
  <dc:subject>Flower database</dc:subject>
  <dc:description>We have created two flower datasets by gathering images from various websites, with some supplementary images from our own photographs. The first dataset is a smaller one consisting of 17 different flower categories, and the second dataset is much larger, consisting of 162 different categories of flowers common to the UK.</dc:description>
  <dc:publisher>University of Oxford</dc:publisher>
</metadata>
```

XML data is valid.

[Buttons: Publish, Cancel]
ODS-FlowerSet1

Path: XU4f1gruQk0kh1UUCZt2kedhLTogqNdpq0jtkWk > FlowerSet1

Public URL: https://onedata.org/share/8abBEfH07aCEBoLxjYdibl8eAgG3v2z...

<table>
<thead>
<tr>
<th>FILES</th>
<th>SIZE</th>
<th>MODIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlowerSet1</td>
<td></td>
<td>2016-09-28 15:09</td>
</tr>
</tbody>
</table>
**Open Data Platform Workflow**

- **Register identifier (e.g. DOI)**
- **Expose share metadata over OAI-PMH**
- **Create a share from folder or file with Metadata in DC**
- **View data set preview in browser using DOI**
- **View data set preview in browser using DOI**
- **Copy data set**
- **Lazy Replication**

**Public Services For Data Discovery**

**Discover data set**

**OpenAIRE**

**Copy of Data-set-1.1**

**Private Resources**

**Share of Dataset**

**Dataset**

**Private Resources**
<table>
<thead>
<tr>
<th>FILES</th>
<th>SIZE</th>
<th>MODIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>_image_0048.jpg</td>
<td>4 KB</td>
<td>2016-09-28 12:09</td>
</tr>
<tr>
<td>_image_0099.jpg</td>
<td>4 KB</td>
<td>2016-09-28 12:09</td>
</tr>
<tr>
<td>image_0001.jpg</td>
<td>50.85 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0002.jpg</td>
<td>41.24 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0003.jpg</td>
<td>46.2 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0004.jpg</td>
<td>30.16 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0005.jpg</td>
<td>32.73 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0006.jpg</td>
<td>27.75 KB</td>
<td>2016-09-28 11:09</td>
</tr>
<tr>
<td>image_0007.jpg</td>
<td>37.93 KB</td>
<td>2016-09-28 11:09</td>
</tr>
</tbody>
</table>
**Open Data Platform Workflow**

- **Register identifier (e.g. DOI)**
- **Expose share metadata over OAI-PMH**
- **Create a share from folder or file with Metadata in DC**
- **View data set preview in browser using DOI**
- **View data set preview in browser using DOI**
- **Copy data set**
- **Copy data set**
- **Lazy Replication**
- **Private Resources**
- **Private Resources**
- **Share of Dataset**
- **Dataset**
- **Public Services For Data Discovery**
- **Discover data set**
- **Exposed share metadata over OAI-PMH**
- **DataCite**
- **OpenAIRE**
PROBLEM 5: METADATA MANAGEMENT INTEGRATED WITH DATA MANAGEMENT PLATFORM

- Work with data and metadata in one system – avoiding problems of consistency
- Monitor metadata data changes through API in order to feed external custom systems
- Advanced data discovery capabilities based on metadata
INTEGRATED METADATA MANAGEMENT

- All files and directories can have a custom user metadata
- API for metadata management
- API for data discovery based on metadata
- Virtual Folders based on metadata tags
- Metadata formats: key-value, JSON, RDF
• Control who knows about your data
• Control who can access data on a single file level
AUTHENTICATION AND AUTHORIZATION

- Pluggable methods of authentication per zone
- Multiple levels of access control
- ACL on files and directories
- Group management
- Token based authentication (macaroons)
- X.509 in prep.
AUTENTICATION AND AUTHORIZATION

- Pluggable methods of authentication per zone
- Multiple levels of access control
- ACL on files and directories
- Group management
- Token based authentication (macaroons)
- X.509 in prep.
AUTHENTICATION AND AUTHORIZATION

- Pluggable methods of authentication per zone
- Multiple levels of access control
- ACL on files and directories
- Group management
- Token based authentication (macaroons)
- X.509 in prep.
PROBLEM 7: EASY INTEGRATION USING API WITH EXTERNAL TOOLS

- Integrate external tools using rich API interfaces with data management platform and build more complex environments for data processing
RICH COLLECTION OF APIs

- APIs for all operations
- Flexible permission checking for APIs
- APIs for full eventually consistent integration with external systems
- API fully described using Swagger for generation of clients based on API specification
- Easy to use simple command line clients for REST API
PROBLEM 8: HIGH-THROUGHPUT PROCESSING

Parallel Processing Nodes using POSIX oneclient, CDMI or REST

Protocols CDMI
Protocols S3
Protocols POSIX VFS

Control, Remote Data Access
CDMI API

Direct Access if possible

Storage Access

Ceph   S3   SWIFT   Lustre
THROUGHPUT TESTS

55Gbit/s
On single node
5 parallel streams
**High-throughput Transfers**

Transfer started by:
- User in GUI
- API-s
- Policy
- Access to Rmt. Data

Block-based transfer:
- Remote Data Access on the fly
- Pre-staging
- Data Migration
- Data Replication

Distributed Priority Queue
For cluster to cluster transfers

WAN
EXAMPLE USECASE
MULTI-CLOUD EARTH OBSERVATION IMAGE PROCESSING PoC

JOB QUEUE

JOBS

LOCATION REASONING

JOB CREATION

EOProc SatCat PROCESSOR

FUSE Client Oneclient

APPLICATION COMPOSITION AND DEPLOYMENT WITH DOCKER COMPOSE

JOB DEPLOYMENT

VM MNG. API

CLOUD ORCHESTRATOR

EO-PoC Spaces

Sentinel 1

Sentinel 2

Results

CESNET

Cyfronet

IPT Poland

AWS

Oneprovider

Oneprovider

Oneprovider

Oneprovider

Parts of Sentinel 1 & 2 Data

Parts of Sentinel 1 & 2 Data

Parts of Sentinel 1 & 2 Data

Parts of Sentinel 1 & 2 Data

DATA LOCATION API

DATA REPLICATION, PRE-STAGING, ON DEMAND TRANSFER
FUTURE PLANS
TOPICS ON ROADMAP

• Further improvements in GUI interface:
  • Visual data transfers
  • Monitoring
  • Storage support promises
  • Custom limited tokens for clients and APIs
• QoS and data management plans
• Data discovery through Elasticsearch
• Cross-zone integration (Kademia DHT)
• Docker volume plugin
• onedata-cli for platform management from command line
• Windows and OSX clients
• Dropbox-like Space sync
• Exposing external data as Onedata Space
QUESTIONS?

Please visit:
www.onedata.org

Michał Orzechowski (CYFRONET AGH)