

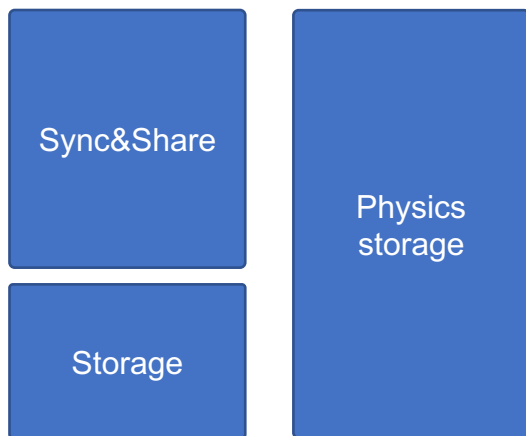
CERNBox, the Scientific Cloud powered by EOS

Diogo Castro
On behalf of the CERNBox team

27 April 2023
EOS 2023 Workshop



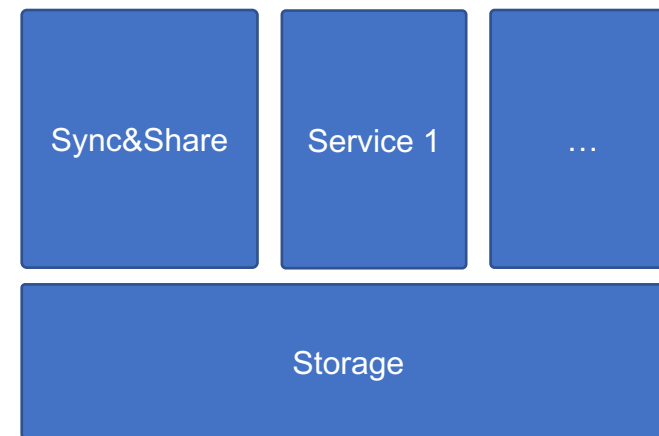
Introduction



Scenario 1



Scenario 2



Scenario 3

- In many sites, physics data and sync&share are split (or each service with its own storage backend)
- There is a trend and will to unify (and user feedback)



What is CERNBox?

Built with **open source** components

Turnkey access to
data



+



+

Safely stored in
CERN campus





CERN(Box) user community

- Extremely diverse
 - Around **37K accounts**
 - Different work backgrounds
 - Different needs
- Different expectations on how the system should support their daily work



Physicists

Service and
Administration

IT

Engineers

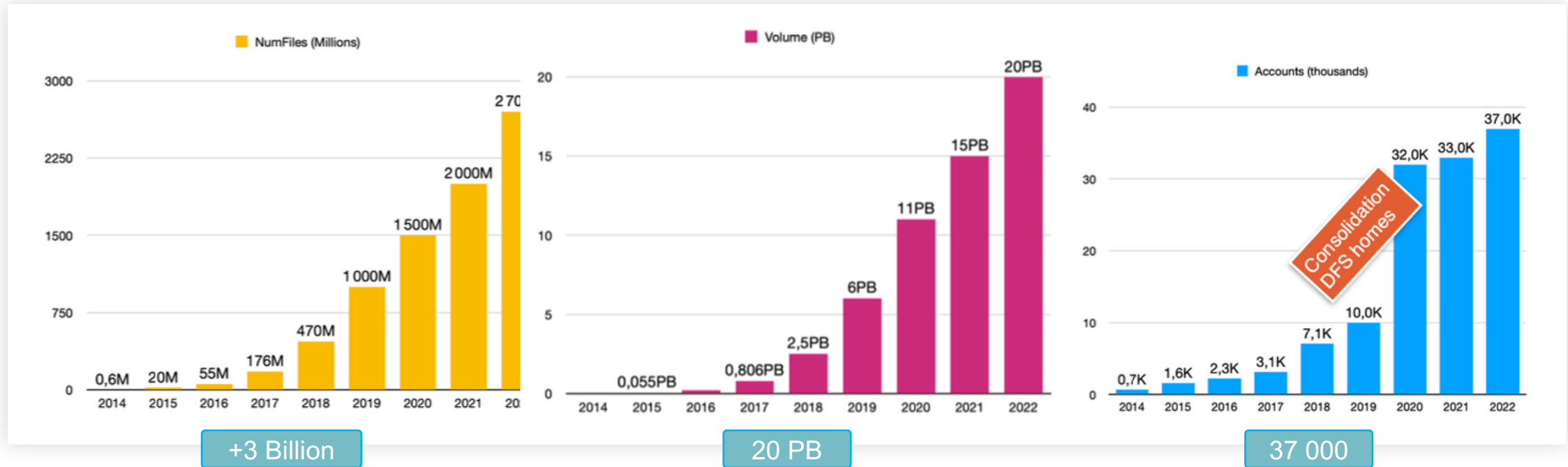


Growing since 2013

- +1000 Project areas
- 31 LHC and non-LHC experiments working areas

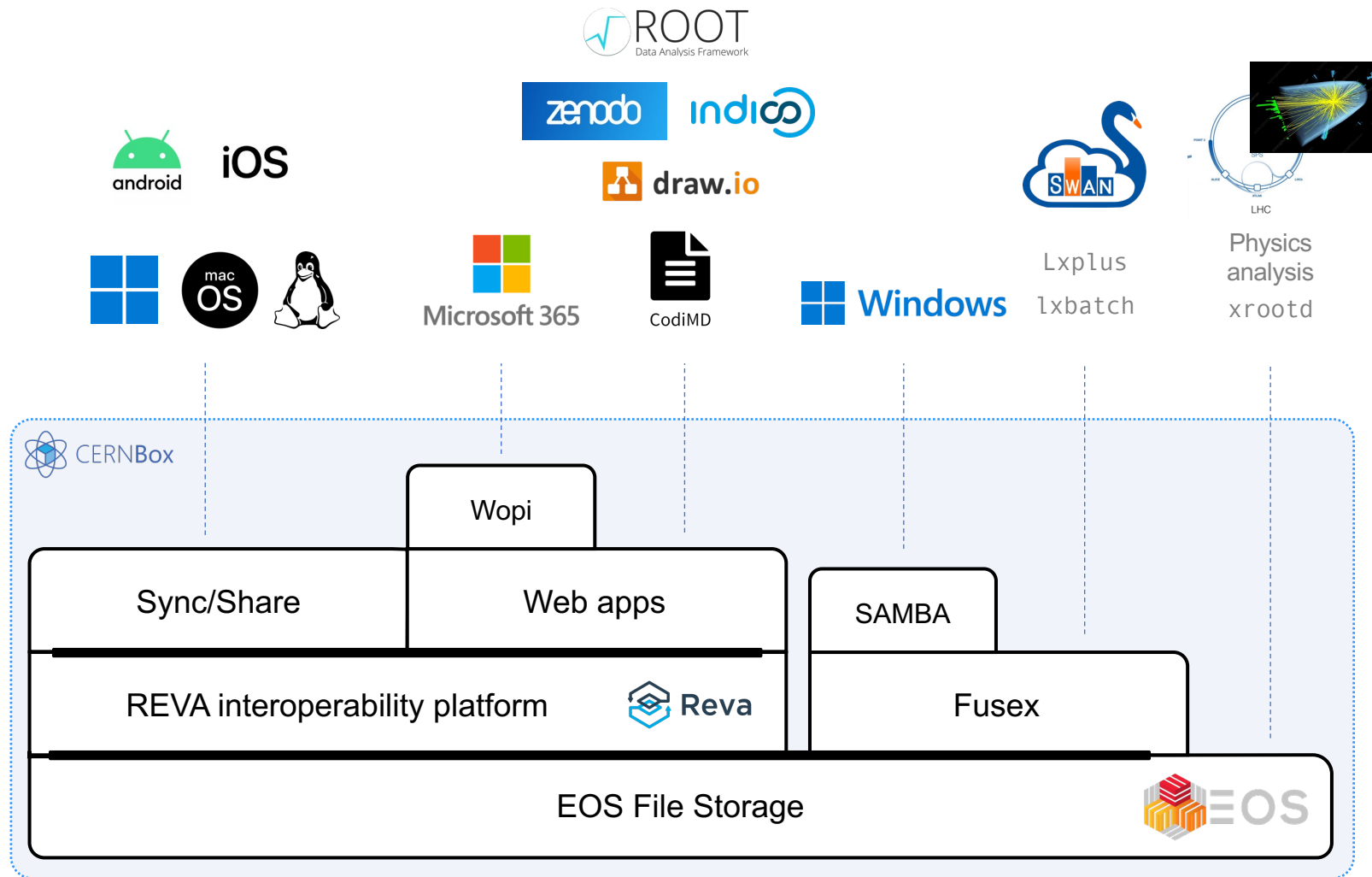
13k unique users/month

Skyrocketing usage





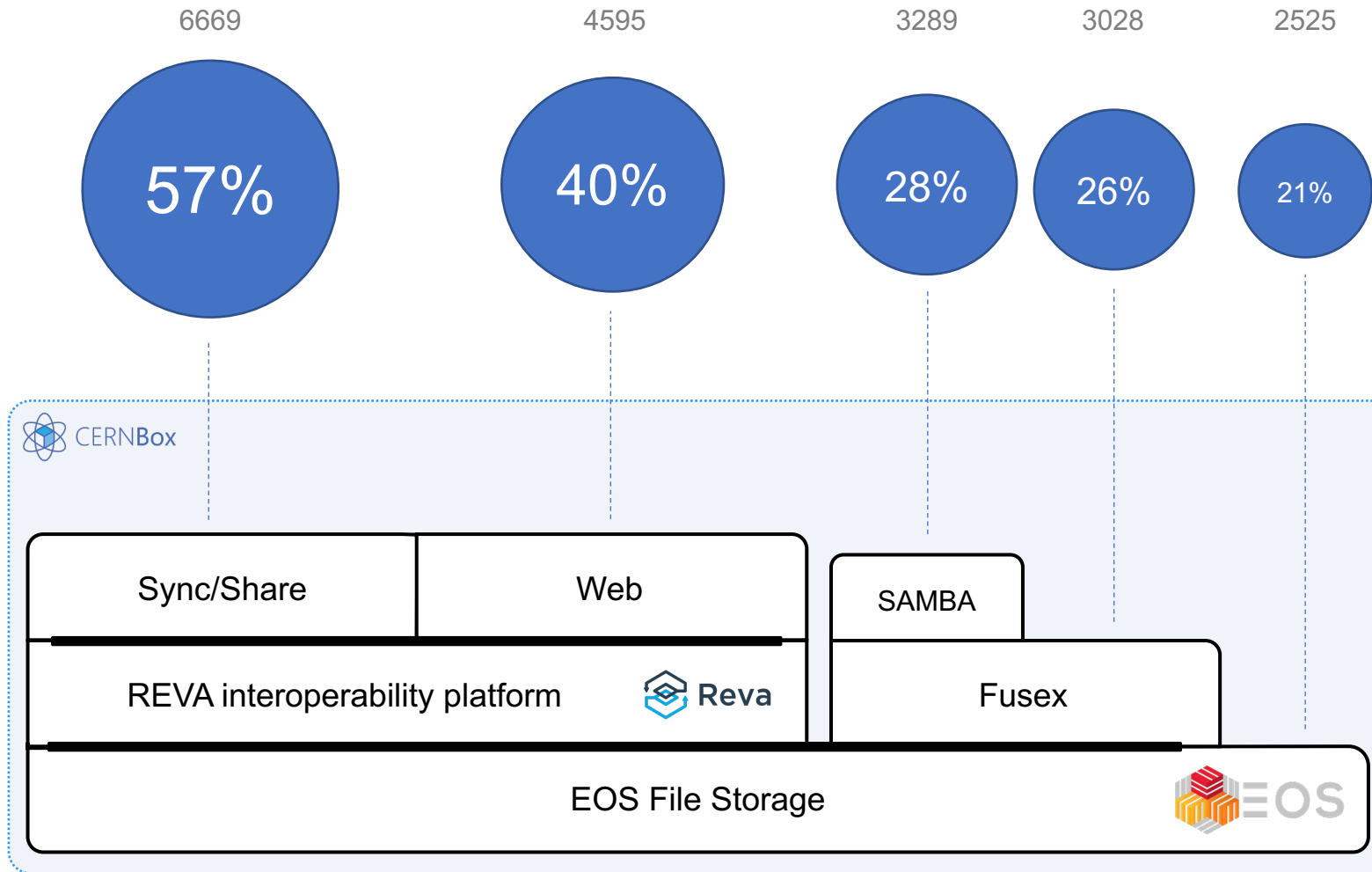
Global architecture



- 37k User accounts
- Sync clients for all major platforms (VFS support on Windows)
- 30k fuse clients
- 15 production clusters
- 18 PB general storage
- 400 PB Physics storage
- <https://reva.link/>



Users per protocol



- Data from 1-14 April
- Unique users across all protocols: **11672**
- Although not the most representative, FUSE and XRoot have the most data transferred in volume



Sharing across protocol boundaries

ACLs are respected independently of the protocol
There's a consistent view everywhere

The collage illustrates the consistency of the file view across different protocols and platforms. The Windows File Explorer shows a local path `H:\user\d\dalvesde\my share`. The web browser view shows the same share on CERNBox. The macOS Finder window shows the same share. The terminal window shows the local file system structure, including folders like `AFS`, `CERNBox`, `Contacts`, `Documents`, `k8s`, `Music`, `Pictures`, `ST`, `SWAN`, `SWAN_projects`, `Test`, and `Videos`.

The dialog box for selecting a remote destination folder shows a tree view of the CERNBox structure, including folders like `eos`, `alice`, `atlas`, `cms`, `engineering`, `experiment`, `lhcb`, and `media`. The path `/eos/user/d/dalvesde/my share` is highlighted.



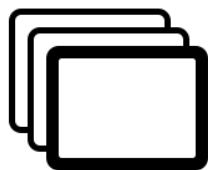
Features, apps and integrations

Features



Backups

Versioning



Sync client and
Mobile apps

Apps & Collaborative editing



Office 365



IFC.js



CodiMD



ROOT
Data Analysis Framework



jupyter



draw.io

Integrations



To infinite scale, and beyond

The new CERNBox version



Challenges: from Sync&Share to Integrated Scientific Platform

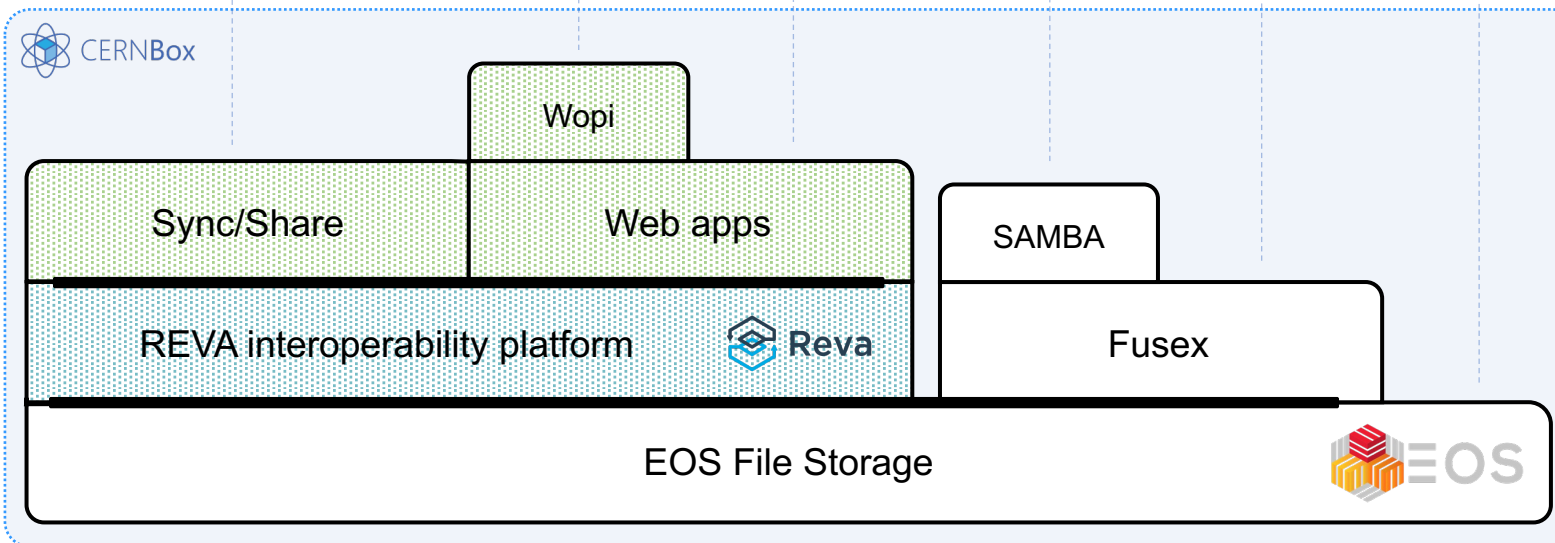
- **Direct Access to Underlying Storage**
 - Assumption that OC is the single entry point
 - Caching might cause inconsistent views and conflicts
- **Homogeneous Access to Shared Data**
 - Need to sync ACLs to storage
- **Data Ownership**
 - Owner of the data is the user running the server (apache/www-data)
 - To increase security and give users full control of their data, the owner should be the real user
- **Native File Versioning and native Recycle bin**
 - Some storage backend provide these, so the OC layer duplicates the functionality
- **Redundant Expensive Calls and inefficient design**
 - Unique filenames for shares requires checking all existing shares before creating a new one



A new solution was required!



The new CERNBox



- New system was deployed to everyone in October 2022
- Based on REVA (<https://reva.link/>) and the new ownCloud Infinite Scale (<https://owncloud.com/infinite-scale/>)



New features

Universal URLs

Native locking

Improved app and projects' integration

Single file sharing

Guest account sharing

Denial ACLs

Indico integration



Add materials to the event.

You can attach files or links using the buttons on the right.



Upload files



Add link



From CERNBox

New UI

Ongoing and future work



New UI features



Self restore
Backup



Search



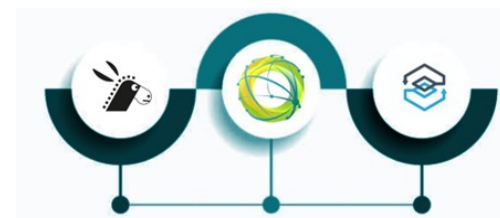
Notifications



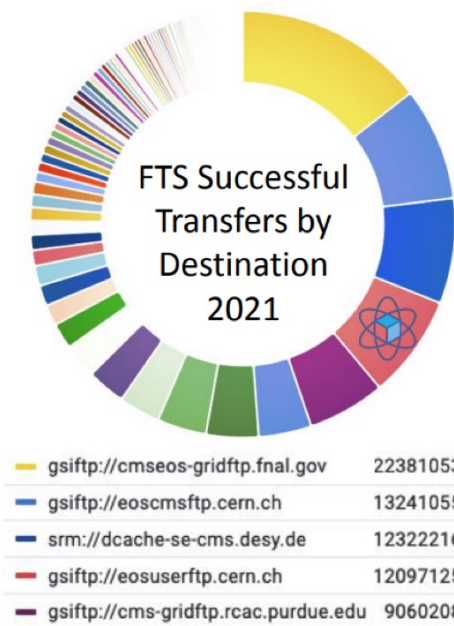
Improved external integrations



- Jupyter native integration with Reva (CS3 APIs)
- Deployment in SWAN



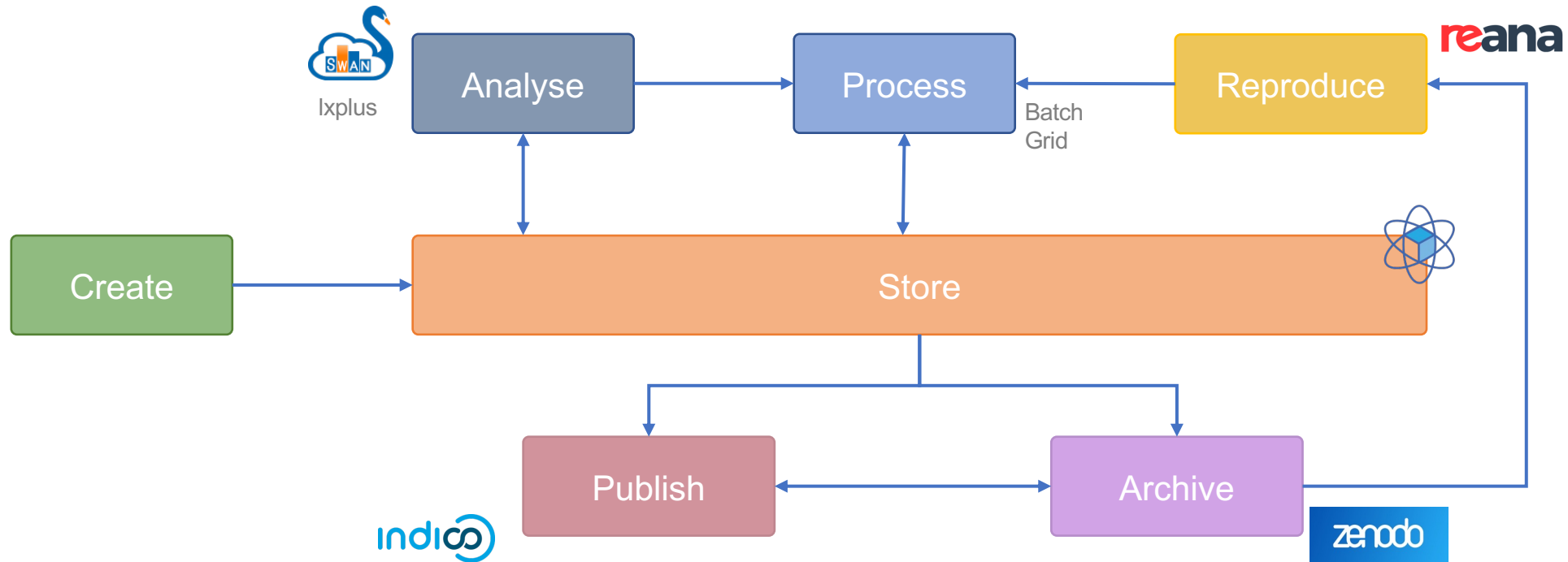
- Integration with FTS
- Native Support for HTTP Third-Party copy





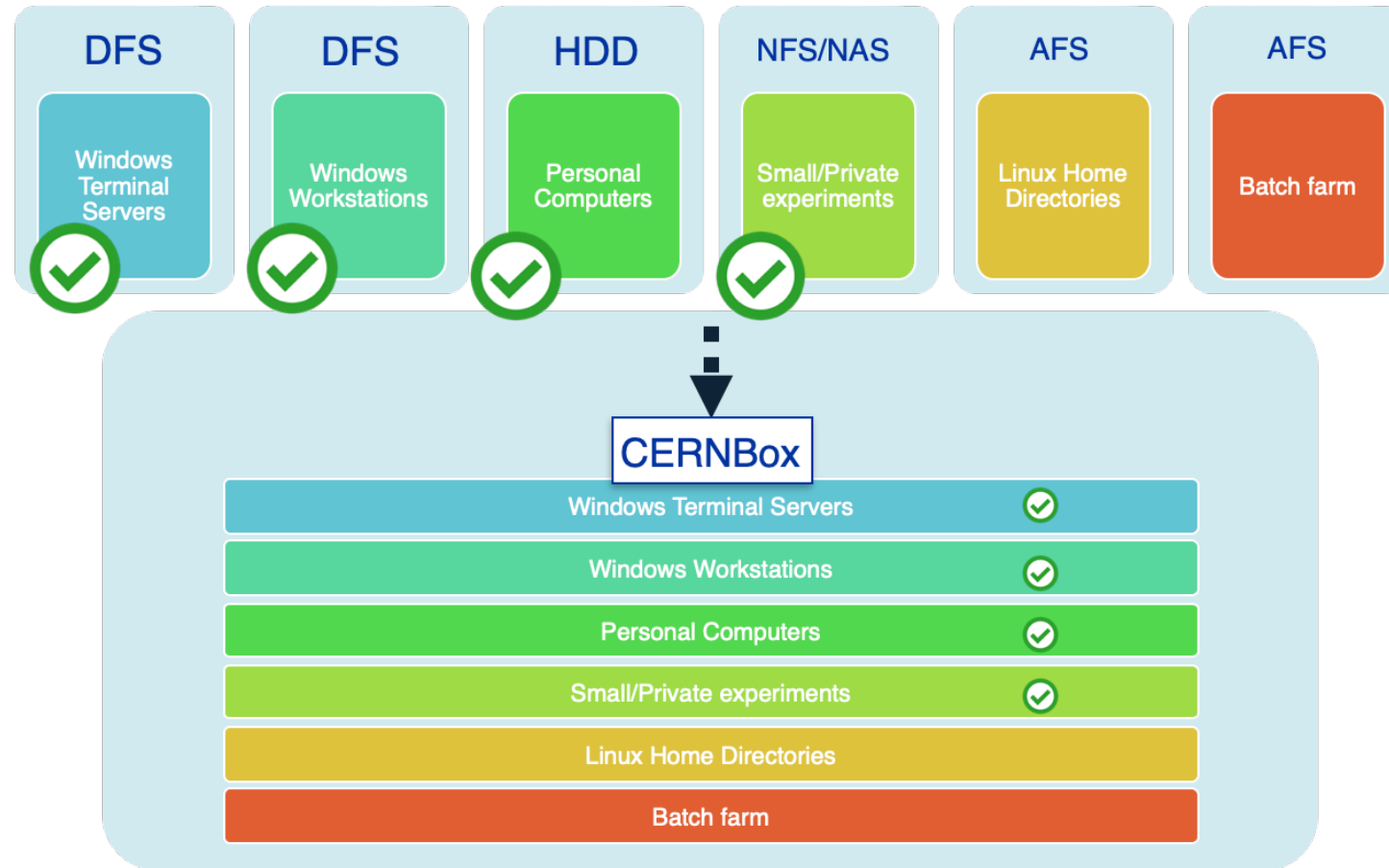
Improved external integrations

- Integration of ScieboRDS for preservation in a Digital Repository (Zenodo/OSF)
 - Supporting Research Data Lifecycle





Consolidation of services





CERNBox: federation of storages using REVA

The image displays the CERNBox web interface, which is a federation of storage systems using REVA. The interface is divided into several sections:

- EOS Atlas:** This section shows the 'EOS Atlas' storage federation. It includes a sidebar with navigation options like 'All files', 'Favorites', 'Shares', 'Spaces', and 'Projects'. The main content area shows a list of folders: 'atlas', 'media', 'project', and 'user'. There are buttons for '+ New' and 'Upload'.
- EOS Media:** This section shows the 'EOS Media' storage federation. It displays a grid of media items, including images and videos, with titles like 'Z' → tt with MZ' = 3000 GeV, for 2016 ATLAS open data release' and 'tt → Jets, for 2016 ATLAS open data release'.
- EOS Public:** This section shows the 'EOS Public' storage federation. It displays a grid of public data items, including images and videos, with titles like 'Z' → tt with MZ' = 3000 GeV, for 2016 ATLAS open data release' and 'tt → Jets, for 2016 ATLAS open data release'.

The interface also features a search bar at the top, a 'What's new?' button, and a 'Light mode' toggle. The user's profile 'John' is visible in the top right corner.



ScienceMesh

- Federated + trusted network of Sync&Share services
 - 300K user base (CS3 community) across Europe (SURF, DESY, PIC, SUNET, ...)
 - Product of EC H2020 funded project CS3MESH4EOSC (ending in June 2023)
 - REVA provides access to the mesh



OPENCLOUD**MESH**



Federated sharing of data across different sync-and-share services.



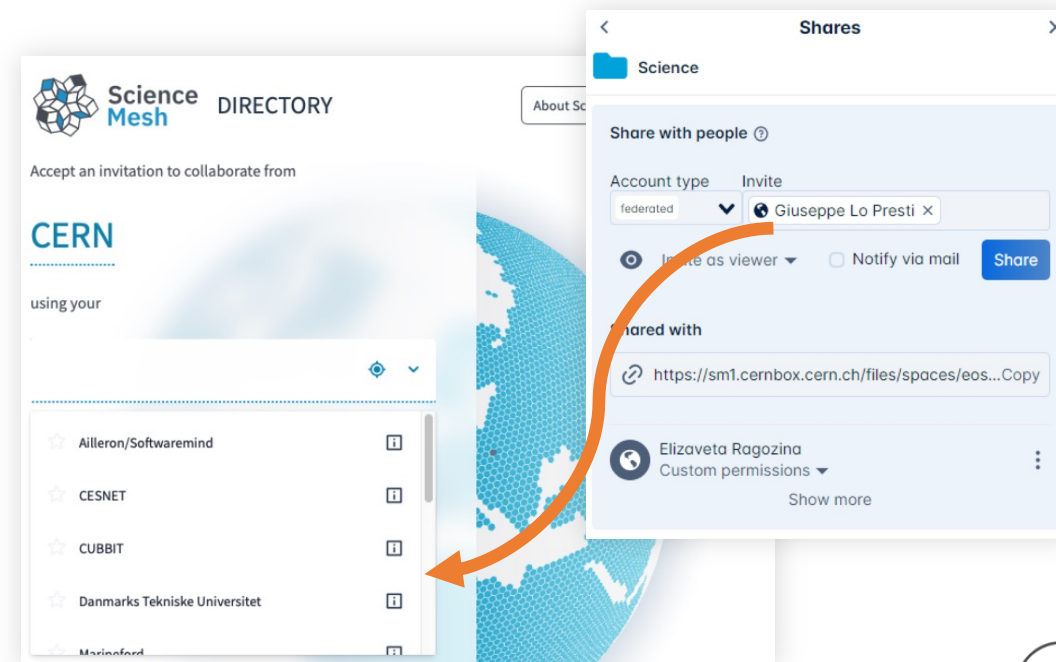
Federated use of applications such as collaborative document editing, data archiving, and data publishing.



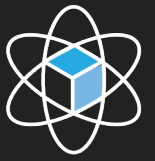
Fast transfer of large datasets from one site to another.



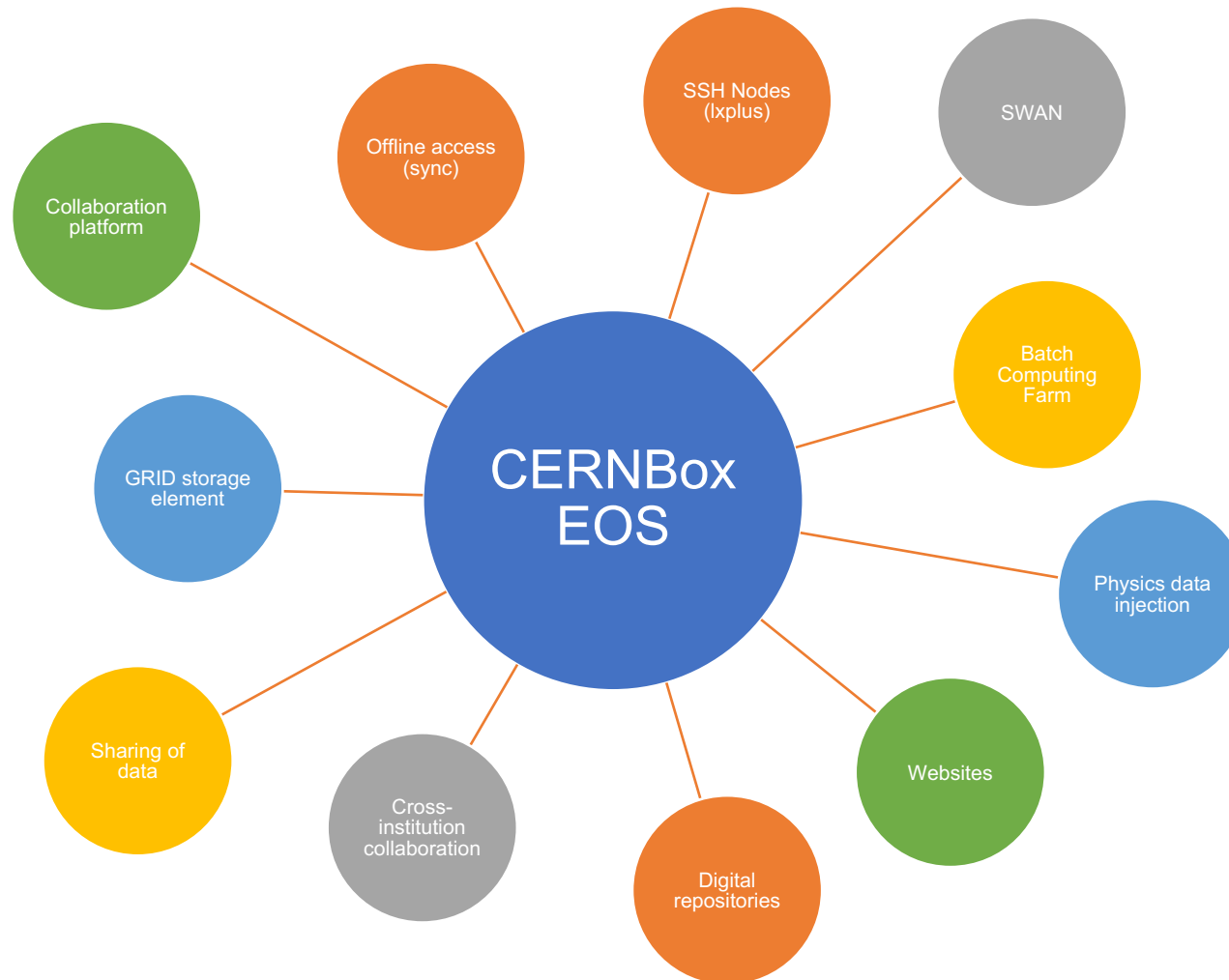
Remote data analysis through sync-and-share services.



Conclusion



CERNBox, the integrated layer for Scientific Data Lifecycle



Reminder:
BoF session this afternoon
on how to deploy a *CERNBox-like* system



CERNBox, the Scientific Cloud powered by EOS

Thank you

Diogo Castro
diogo.castro@cern.ch