LABDRIVE
Research Data Management
Platform training
<table>
<thead>
<tr>
<th>Agenda</th>
<th>Duration</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABDRIVE introduction</td>
<td>60 mins</td>
<td>9:30-10:20</td>
</tr>
<tr>
<td>Architecture and overview</td>
<td>35 mins</td>
<td>9:30-10:05</td>
</tr>
<tr>
<td>How research content is to be organized</td>
<td>25 mins</td>
<td>10:05-10:30</td>
</tr>
<tr>
<td><strong>Break</strong></td>
<td></td>
<td>10:30-10:45</td>
</tr>
<tr>
<td>Configuration</td>
<td>70 mins</td>
<td>10:45-12:10</td>
</tr>
<tr>
<td>Users and Permissions</td>
<td>15 mins</td>
<td>10:45-11:00</td>
</tr>
<tr>
<td>Archival organization</td>
<td>15 mins</td>
<td>11:00-11:10</td>
</tr>
<tr>
<td>Container – concept and usage</td>
<td>10 mins</td>
<td>11:10-11:25</td>
</tr>
<tr>
<td>Introduction to metadata – concept and usage (container, item &amp; tags)</td>
<td>15 mins</td>
<td>11:25-11:40</td>
</tr>
<tr>
<td>Metadata configuration</td>
<td>15 mins</td>
<td>11:40-11:55</td>
</tr>
<tr>
<td><strong>Break</strong></td>
<td></td>
<td>11:55-12:10</td>
</tr>
<tr>
<td>Operations</td>
<td>110 mins</td>
<td>12:10-13:20</td>
</tr>
<tr>
<td>Create a data container</td>
<td>10 mins</td>
<td>12:10-12:20</td>
</tr>
<tr>
<td>Upload content</td>
<td>10 mins</td>
<td>12:20-12:30</td>
</tr>
<tr>
<td>Download content</td>
<td>10 mins</td>
<td>12:30-12:40</td>
</tr>
<tr>
<td>Searching</td>
<td>20 mins</td>
<td>12:40-13:00</td>
</tr>
<tr>
<td><strong>Lunch Break</strong></td>
<td></td>
<td>13:00-13:30</td>
</tr>
<tr>
<td>LABDRIVE functions &amp; workflows</td>
<td>20 mins</td>
<td>13:30-13:50</td>
</tr>
<tr>
<td>Storage mode transitions</td>
<td>10 mins</td>
<td>13:50-14:00</td>
</tr>
<tr>
<td>Reports</td>
<td>10 mins</td>
<td>14:00-14:10</td>
</tr>
<tr>
<td>Advanced operations – Jupyter Notebooks &amp; reproducibility</td>
<td>20 mins</td>
<td>14:10-14:30</td>
</tr>
<tr>
<td><strong>Q&amp;A &amp; Conclusions</strong></td>
<td>15 mins</td>
<td>14:30-14:45</td>
</tr>
</tbody>
</table>
About LIBNOVA
About LIBNOVA

LIBNOVA is market leader in digital preservation and digital content archiving.

Organizations use our solutions to safeguard and provide access to their valuable digital assets for the long term, either by using our cloud platform or by deploying our software on their own premises.

Present in 17 countries. Some of the largest and most demanding organizations worldwide are using our platforms.

Founded in 2009. Modern architecture. Massively adopted during the last 5 years.

Self-sustained company, no risk capital, no debt.

We work with them with a long-term, partnership approach.
LABDRIVE Introduction:
Overview
LABDRIVE is a **Research Data Management** and Preservation platform.

Its design principles are:

- **Unified**: It enables organizations to **combine** their research content in a single, unified, rationalized platform.

- **Comprehensive**: It allows organizations to **capture the research** data they produce, helping them to **properly manage, preserve and allow access to it**, during the whole data lifecycle (and not only at the end of the cycle).

- **Open**: ISO16363, ISO27001, ISO27017, ISO27018 certified. 100% of the information can be easily exported. No exit barriers. User-extensible.
LABDRIVE allows organizations to **organize and unify** their content:

Transition from a **siloed approach** in which each series of datasets, experiments, departments or units are using multiple, disaggregated systems to keep content

To a **unified repository** that can adapt to the particularities of each dataset, **unifying all content in a single platform**.

- Risk management
- Publishing/rights management
- Permissions control
- Policies management
- Infrastructure management
- Cost control
LABDRIVE is a comprehensive solution:

- **Management**, not only archiving: It allows organizations to **capture the research** data they produce, helping them to **properly manage, preserve and allow access to it**, during the whole data lifecycle (and not only at the end of the cycle).
LABDRIVE is a comprehensive solution:

- **Unique approach to package generation:** Packages don’t need to be fully defined for ingestion. They can be enriched and modified over time. LABDRIVE manages the complexity.

- **Adapts to different needs:** Organizations can create working areas with different access methods, metadata schemas, workflows, permissions, rights, storage and cost.

- **Scales:** 15PB ingestion has been demonstrated (at 500TB/day ingestion rate)
LABDRIVE is based on open standards and has a clear exit path.

- Externally certified alignment to OAIS and ISO16363: ISO16363 is the **gold standard** in digital preservation. LABDRIVE is getting certified in July 2022. Platform allows to create fully supports the OAIS Information model.

- Fully aligned with the TRUST and FAIR principles.

- The platform is certified and externally audited in some of the most demanding security and compliance standards: ISO27001, ISO27017, ISO27018.

- 100% of the information can be retrieved at any time. Can run in the cloud or on-premises.
LABDRIVE is a **Research Data Management and Preservation** platform.

Its design principles are:

- **Unified**: It enables organizations to **combine** their research content in a single, unified, rationalized platform.

- **Comprehensive**: It allows organizations to **capture the research** data they produce, helping them to **properly manage, preserve and allow access to it**, during the whole data lifecycle (and not only at the end of the cycle).

- **Open**: ISO16363, ISO27001, ISO27017, ISO27018 certified. 100% of the information can be easily exported. No exit barriers.
LABDRIVE Introduction:
Architecture
Architecture

Web interface + API ingress
Protocol Servers (web share, S3, NFS, XrootD, etc.)
JUPYTER Kernels (isolated)
Functions Engine (NUCLIO)
File validation Agents
Property extractor (hashes, full text, characterization metadata, …)
Actions agents
Infrastructure migration and integrity assurance agents

AWS S3 Storage / CEPH
Storage

Kubernetes
Database service
Elasticsearch service
Infrastructure

LABDRIVE
• Users are able to talk directly and transparently with the AWS S3 storage
  -> The platform becomes massively scalable (up/downloads)
  -> Massively interoperable. Anything that works with S3 can be used to preserve or to interact with
  preserved data: scripting, SDKs, etc. but also Analytics tools, reporting, reproducibility environments,
  etc. Thousands of components can be easily connected.
LABDRIVE Introduction:
How research content is to be organized
LABDRIVE offers a great degree of flexibility on how information is organized when in the platform, and it is able to adapt to any kind of data structure.

In LABDRIVE, there is no imposed data structure: organizations and users are in charge of defining their own data models for their content.

What are the tools LABDRIVE provide for you to do so?
- Archival structure/nodes
- Data containers
- File/folder structures
- Per-item metadata
- Workflows (Functions)
- Compliance reports
How research content is to be organized

Archival structure/nodes. Data containers. File/folder structures and Per-item metadata

Unit, department, teams, groups, content types...

Objects, experiments, packages, files....
How research content is to be organized

**Workflows**

- **Default workflow**
  - Unique ID generation
  - Integrity (hashes) generation
  - Structural information
  - Properties extraction
  - Malware scan
  - Characterization
  - Packaging
  - Storage

- **Extended workflows (optional)**
  - Fine with default (No extended workflow)
  - Or things like:
    - Extract TAR then
    - Validate BagIt contents then
    - Verify integrity then
    - Load technical metadata from JSON then
    - Create report
How research content is to be organized

Compliance reports

- In general, LABDRIVE adopts the simplicity and flexibility as paradigms.

- For example:
  - Instead of making a metadata field mandatory, we aim at creating a metadata compliance report, that highlights when mandatory fields are not there.
  - Instead of forbidding to upload a certain file format, we aim at using a file format report, that highlights the file formats that should not be used.
  - Instead of forbidding the upload of malware, the platform reports it.
LABDRIVE Configuration
Configuration

▪ Users and permissions
  ▪ Groups
  ▪ Application permissions
  ▪ Content permissions
  ▪ Federated auth
    ▪ Automatic account creation
    ▪ Automated account configuration
  ▪ Permission audit:
    ▪ Effective permissions
    ▪ Users
  ▪ Security audit
Permissions

**BY APPLICATION**

- Containers section
  - View disabled containers
- Reports section
  - Used space by Archival Structure
  - Used space by Containers
  - Pools
  - Storage use per archival node
  - Storage use per Container
  - Effective permissions audit
  - Security audit
  - Users
- Configuration section
  - Archival structure section
  - Submission areas section
  - Content tag section
  - Workflow section
  - Lifecycle policies section
  - File formats section
  - Container metadata section
  - Object metadata section
  - Users section
  - Groups section
  - Data Container Templates section
  - Functions section
  - Federated Authentication section
  - Access methods section
  - Manual section
  - AIP Disc section

**BY NODE**

- Create Container
  - User is able to create new containers.
- Read Container
  - User is able to see the container (but not necessarily its content).
- Update Container
  - User can change the container metadata, details, workflow, quota and other Container related settings.
- Delete Container
  - User can archive, permanently delete the container (and all its files or folders), associated metadata, events, versions and all related elements. This permission also allows the user to archive or un-archive deleted containers.
- Read Container Content
  - User can see, but not change, files and metadata inside the container. Versions for files, if they exist, are also visible (but cannot recover older versions of a file).
- Write Container Content
  - User can create, overwrite and delete files, but not to modify the file metadata or versions for files, if they exist, can be recovered (but not deleted).
- Write Container Content Metadata
  - User can also create, delete or change file/folders metadata.
- Manage Container Storage
  - User is able to define the initial storage class for the container, but is also able to modify it to a transition objects from one storage type to another. As transitioning from one storage status to another usually clears versions and deleted items, this user is also capable of permanently storing deleted items and object versions.
- Purge Deleted Items
  - User is able to permanently delete all already deleted items and other versions of existing files.

**BY CONTAINER**

- Read Container
  - User can see the Container (but not necessarily its content).
- Read Container Log
  - User is able to see its own and other users actions in the event.
- Run Core Functions
  - User is able to run built-in core functions like bulk metadata edit, move content to other containers, etc.
- Run User Functions
  - User is able to run run-time defined user functions
- Node Admin
  - User is able to create new sub nodes, assign and change permissions and ultimately perform any operation.

**Permissions**

- Permissions write
  - User can see and modify the permissions associated to the container (even for itself). If you grant this permission to a user, the user is able to modify the permissions for everything else (so may be virtually have any permission)
- Permissions delete
  - User is able to add new users or groups to the permissions for the container.
- Read Container Log
  - User is able to see its own and other users actions in the event.
- Run Core Functions
  - User is able to run built-in core functions like bulk metadata edit, move content to other containers, etc.
- Run User Functions
  - User is able to run run-time defined user functions
- Node Admin
  - User is able to create new sub nodes, assign and change permissions and ultimately perform any operation.

**Delete Container**

- User can archive, permanently delete the container (and all its files or folders), associated metadata, events, versions and all related elements. This permission also allows the user to archive or un-archive deleted containers.

**Write Container Content**

- User can create, overwrite and delete files, but not to modify the file metadata or versions for files, if they exist, can be recovered (but not deleted).

**Purge Deleted Items**

- User is able to permanently delete all already deleted items and other versions of existing files.

**Write Container Content Metadata**

- User can create, overwrite and delete files, but not to modify the file metadata or versions for files, if they exist, can be recovered (but not deleted).

**Manage Container Storage**

- User is able to define the initial storage class for the container, but is also able to modify it to a transition objects from one storage type to another. As transitioning from one storage status to another usually clears versions and deleted items, this user is also capable of permanently storing deleted items and object versions.

**Purge Deleted Items**

- User is able to permanently delete all already deleted items and other versions of existing files.

**Write Container Content Metadata**

- User can create, overwrite and delete files, but not to modify the file metadata or versions for files, if they exist, can be recovered (but not deleted).

**Permissions write**

- User can see and modify the permissions associated to the container (even for itself). If you grant this permission to a user, the user is able to modify the permissions for everything else (so may be virtually have any permission)

**Permissions delete**

- User is able to add new users or groups to the permissions for the container.

**Read Container Log**

- User is able to see its own and other users actions in the event.

**Run Core Functions**

- User is able to run built-in core functions like bulk metadata edit, move content to other containers, etc.

**Run User Functions**

- User is able to run run-time defined user functions

**Node Admin**

- User is able to create new sub nodes, assign and change permissions and ultimately perform any operation.
When creating sub-nodes or containers, the user can choose between inheriting the parent permissions or customise. Customised permissions override the permissions at a previous level (whether it is a root node, or a sub-node).

User and groups permissions add to each other when combined. Therefore, if user A is given individual permissions and then assigned to a group with its own permissions, user A retains their individual permissions and adds the new ones given by the group, regardless of their old ones.
The user or group has the right to see and manipulate containers A and B as allowed by the defined parent node permissions. However, custom permissions have been selected for container C and left blank, which leaves the user unable to see nor manipulate the container.
Configuration

- Archival organization
  - Nodes
  - Node templates

- Containers
  - Concepts
    - Overview
    - Metadata
    - Storage

- Introduction to metadata and metadata configuration
  - How it works
  - Schema
  - Importing/exporting
Operation

- Create a container
  - GUI
  - API
- Uploading content
  - Browser
  - S3
- Downloading content
  - Browser
  - S3
- Searching
  - GUI
  - API
- Functions and Workflows
  - Overview
  - Sample
  - API
- Storage
(Set your environment variables: API key + S3)
(Offsets and limits)

List Archival Structure
Create an Archival Node
List Archival Node permissions
Edit Archival Node permissions
  List groups
  List permissions
  Assign a permission to a group for an Archival Node

Create a container
  List container metadata schemas
  List object metadata schemas
  List workflows
  Create container
Operation - API

Uploading content

Downloading content
  The S3 way
  The API way
    Listing objects (file/folder)
    Downloading
Operation

- Storage
- Reports
- Jupyter notebooks / snakemake / Custom containers
Thanks!

contact@libnova.com