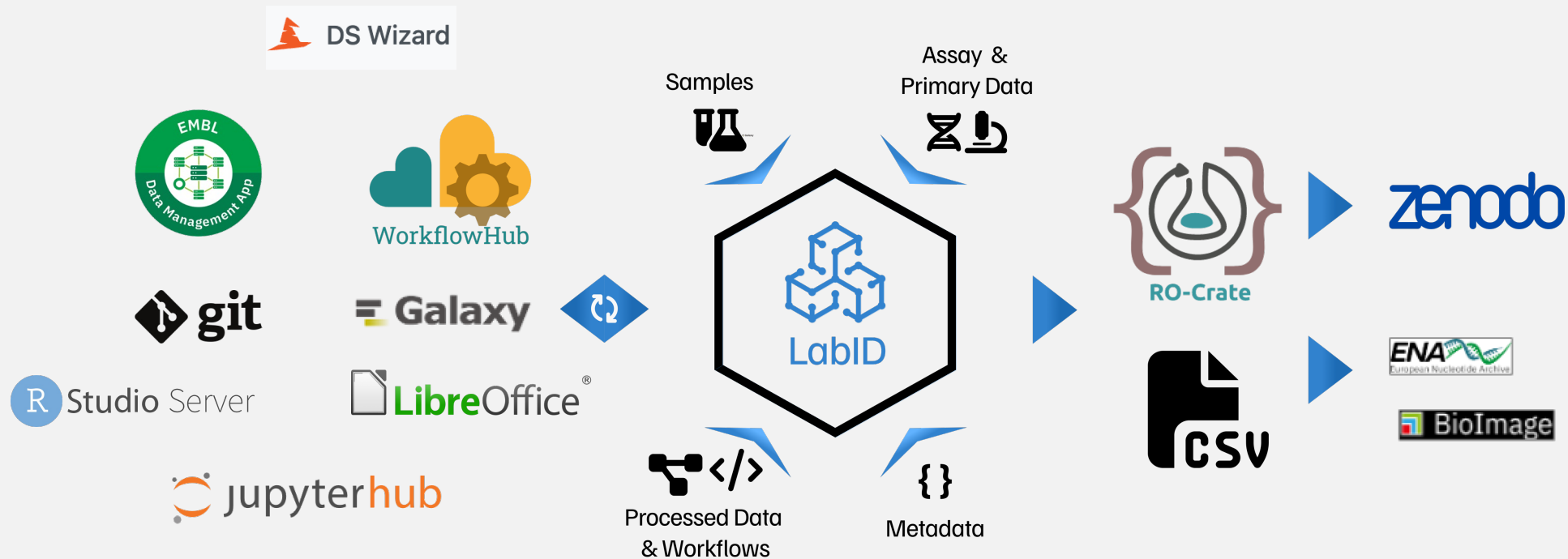


LabID Integration Capabilities and Data Flows



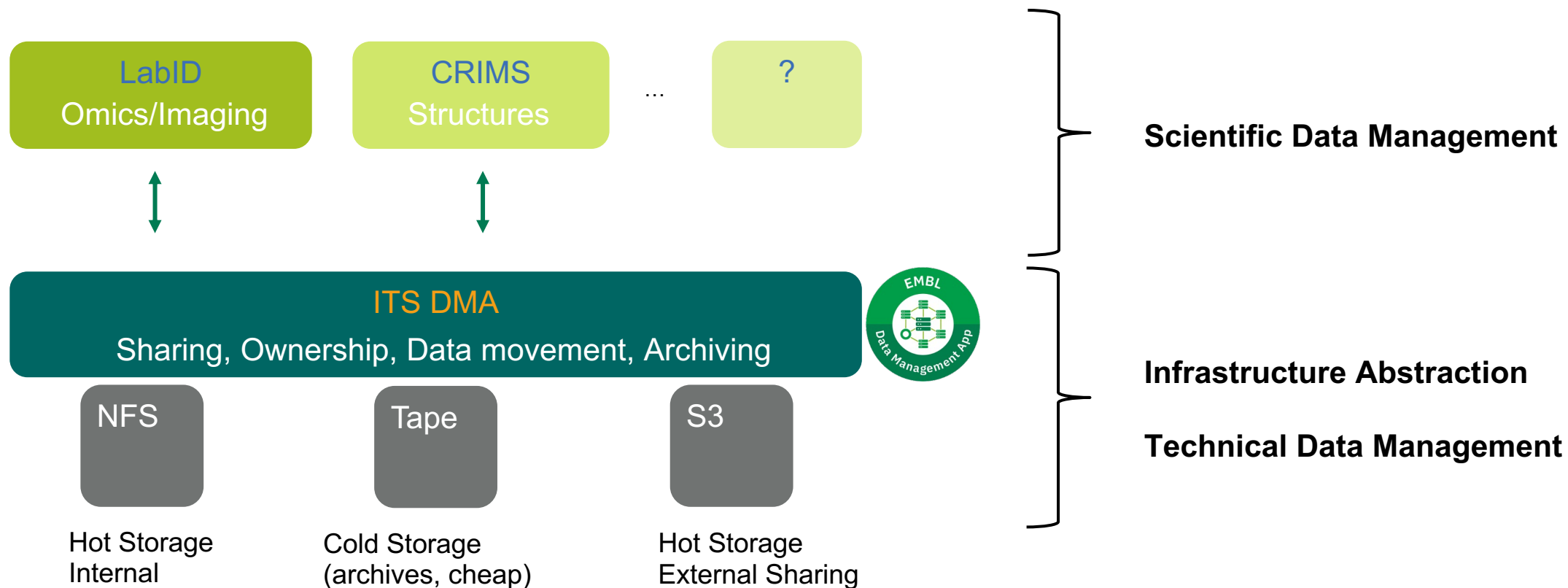
Charles Girardot
Multimodal Open Data Integration Support (MODIS)
Data Science Center

LabID is one piece of a large ecosystem



- LabID does not work in isolation
- Integrate with many other tools through REST APIs and standards

The DM Stack at EMBL



The EMBL ITS **DMA** app

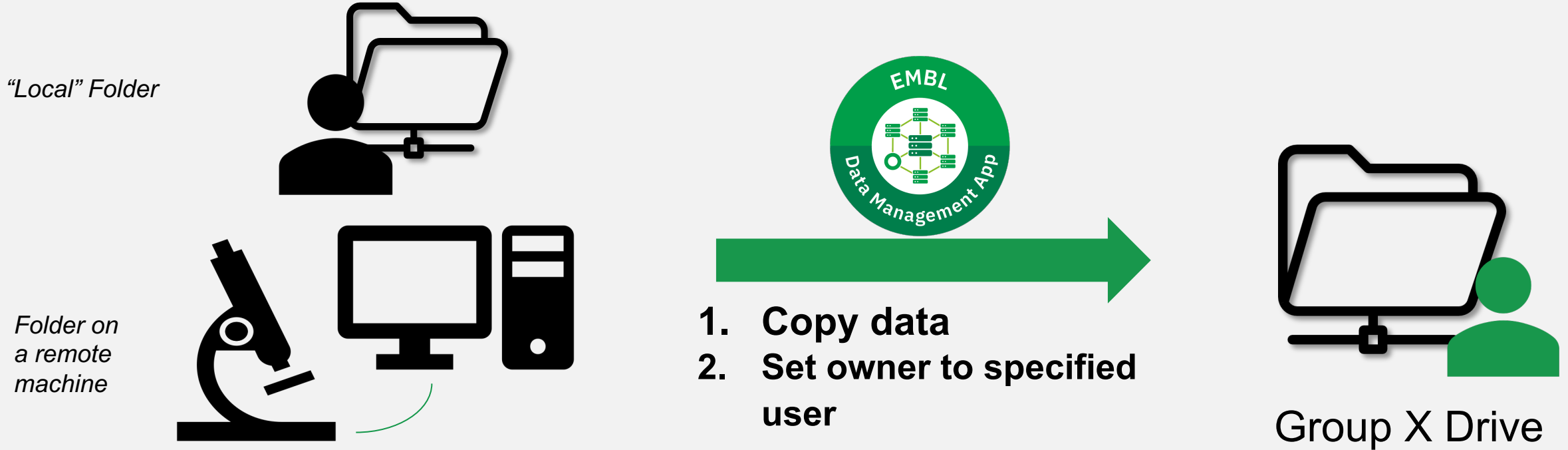
- tracks data's location, ownership & metadata
- allows performing various data operations on the IT infrastructure

=> LabID integrates with DMA to delegate various operations through DMA REST API

The DMA's REST API allows integration with LabID

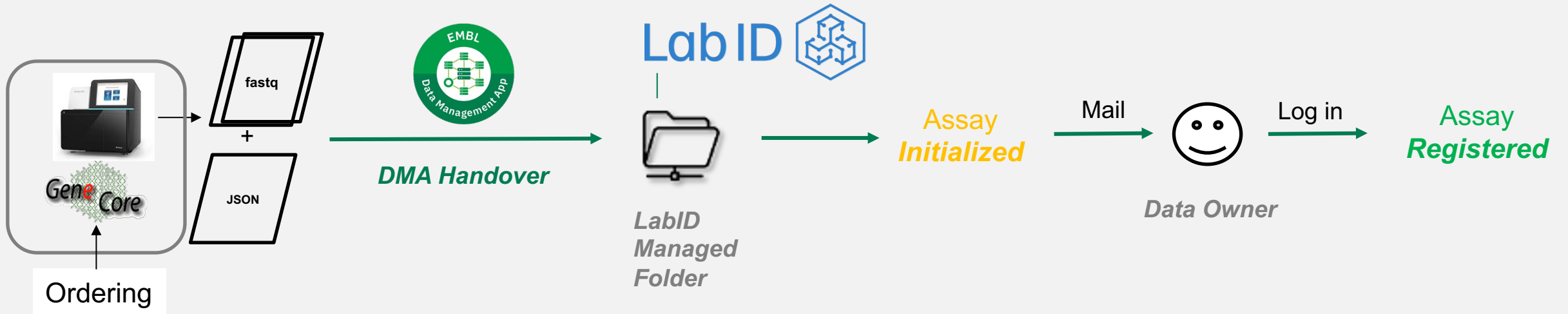


Data Movement – DMA “Handover” & “Pullover”



- Used by several Core Facilities to distribute data
 - Changing ownership requires Elevated Privileges on unix systems
- Used by users to move their data from acquisition machine

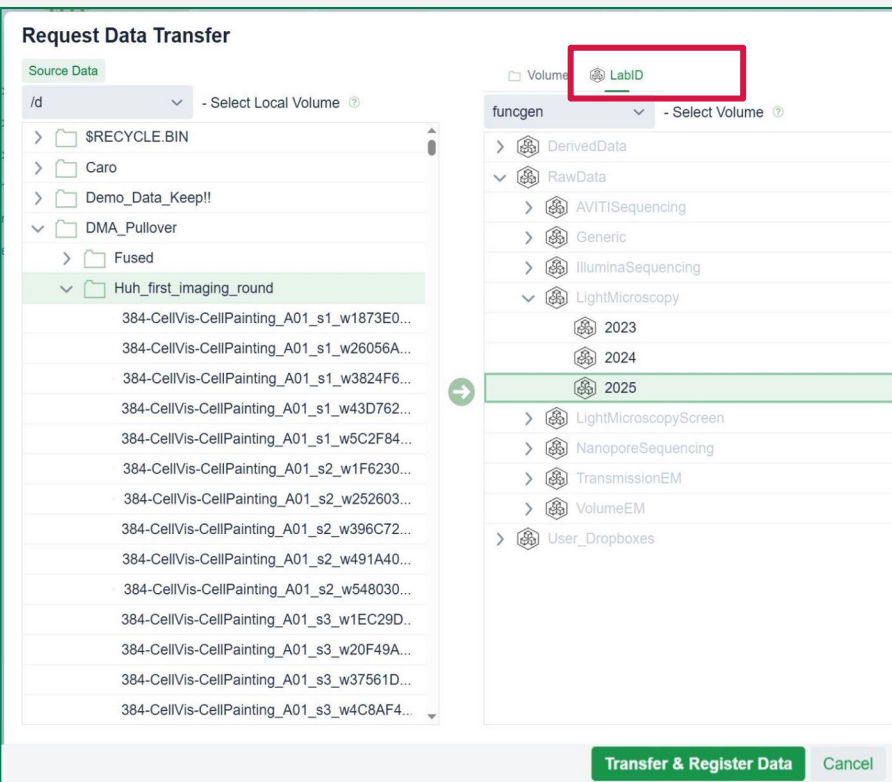
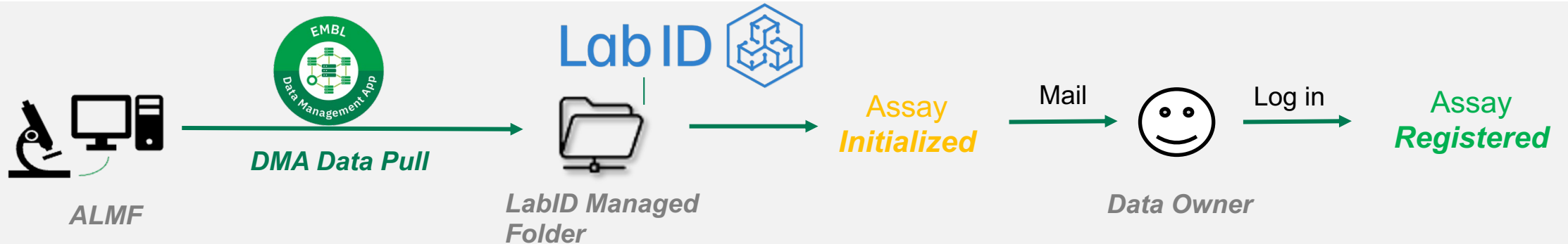
Automated Transfer of Sequencing Data from GeneCore



- Minimum information collection (metadata from GeneCore)
 - simplified data registration wizard
- No data loss: every sequencing run is captured
- Data stored on owner's storage
- Data protected from user manipulation (read-only)
- Data remains accessible for analysis (no duplication needed)

✓ Ideal scenario, users do not handle the data themselves

Streamlining imaging data registration with Data Pullover



- User initiates registration through DMA Pull Over **from the acquisition machine**
- The destination selection provides minimum context (e.g. assay type)
- Image specific metadata available (XYZCT)

Data Registration Automation

One monitored folder



- Automate (primary) data registration in well defined scenarios (e.g. experiment type, project)
- Registration only happens after successful content validation
 - Data structure, Naming convention (as defined in DMP)
 - Rule enforcement e.g. samples must already exist...
- Similar assays are registered the same way
- Metadata Extraction
- Plugin approach (“Sniffer”)

Imaging Session Example

Assay Details

Registered

Name

TREC Plankton Scans ATH 20240705 LSM900

ID

9555601e-5d07-4619-8473-d794f8a83769

State

REGISTERED

Imaging Methods

fluorescence microscopy
confocal microscopy

Instrument

AML-Zeiss LSM 900

Imaging Session

Imaging session ATH 20240705

Objective

—

Run Directory

/g/mobilelab/LabID/Data/Assay/light_microscopy/2024/TREC_STOP_41_Athens_LSM900_20240705

Data Session

2024_12_05_13_21_07.70_wiegand

Description

—

Protocol List

—

XYZCT Dimension Details

Image X Size (px)

2048

Image Y Size (px)

2048

Pixel Size (microns)

0.312

Z Stack

no

Z Planes

1

Z Step (microns)

—

Time Lapse

no

Time Points (rounds)

1

Time Step (s)

—

Visualized Entities per Channel

Channel Index	Name	Target	Label	Probe	Wavelength
1	Hoechst	DNA and surface structures (teca) (DNA)	Hoechst33342 (Fluorophore; Ex/Em max: 352/455 nm)	—	—
2	Chl <i>a</i>	Chlorophyll A autofluorescence (Chloroplast)	—	—	—
3	T-PMT	None (—)	—	—	—
4	PE	Phycocerythrin autofluorescence (—)	—	—	—

Sharing

Ownership and Lifecycle

Annotations

Lineage Graph

Assays (1)

Datasets (8)

Samples (8)

Specimen (8)

Notes

Output Datasets

1 - 8 of 8

Each CZI file is a dataset

Dataset definition is often context specific

Ex: IMATREC, BioOcean5D



Dataset Metadata extraction e.g. for IMATREC

The screenshot shows the 'Item Details' page for a dataset item. The top bar indicates the item is 'Online'. The left sidebar contains a list of metadata fields: ID, Name, Description, Type, Is Raw, QC, Status, Samples, Parent datasets, Children, Assay, Studies, Dataset Collections, Workflow Run, Workflow Version, Workflow, Workflow Data Type, ProtocolList, Archives, Virtual, and Is managed. The main content area displays the details for the selected item, including a JSON representation of the extracted metadata. The JSON is expanded to show the 'Image Extracted Metadata' section, which contains a 'binning' object with 'channels' and 'dimensions' arrays, and a 'scene_number' of 1. The 'dimensions' array lists 'C', 'T', 'X', 'Y', and 'Z' with their respective values. The 'channels' array lists 'TL - Transmitted', 'RF - Reflected', and 'DAPI'.

Item Details Online

ID 6b389ccb-abfe-4306-8de1-66acb5b7ce38

Name BAR_10to40_20240313_PM_01_WM_01_112518454_5x CZI Image

Description CZI image at 5X magnification

Type Generic

Is Raw ✓ yes

QC —

Status ONLINE

Samples BAR_10to40_20240313_PM_01_WM_01

Parent datasets —

Children —

Assay 20250807_IMATREC_CryoLM_WM_5x

Studies IMATREC

Dataset Collections 20250807_IMATREC_CryoLM_WM_5x CZI images

Workflow Run —

Workflow Version —

Workflow —

Workflow Data Type —

ProtocolList —

Archives —

Virtual ✗ no

Is managed ✗ no

Ownership and Lifecycle >

Sharing >

Annotations 14

Pick an annotation type to add

DMAApp 172926 created on January 22, 2026

Image Dimension Order TCZYX

Image Extracted Metadata

```
{
  "binning": "1x1",
  "channels": [
    { "...", // 7 items
    { "...", // 7 items
    { "...", // 7 items
    { "...", // 7 items
    { "...", // 7 items
    { "...", // 7 items
  ],
  "imaged_by": "brey",
  "objective": "5x/0.2Air",
  "dimensions": {
    "C": 6,
    "T": 1,
    "X": 1920,
    "Y": 1216,
    "Z": 1,
    "order": "TCZYX"
  },
  "scene_number": 1,
  "channel_names": [
    "TL - Transmitted",
    "RF - Reflected",
    "DAPI"
  ]
}
```

Magnification 5X

Ordered Channel Names TL - Transmitted; RF - Reflected; Ex353; Ex493; Ex577; Ex653

Physical Pixel Size X (µm) 1.8603174603174604

Physical Pixel Size Y (µm) 1.8603174603174604

Physical Pixel Size Z (µm) None

Image Channel Number 6

Image Time Points 1

Image Size X 1920

Image Size Y 1216

Image Size Z 1

Time Increment (second) None

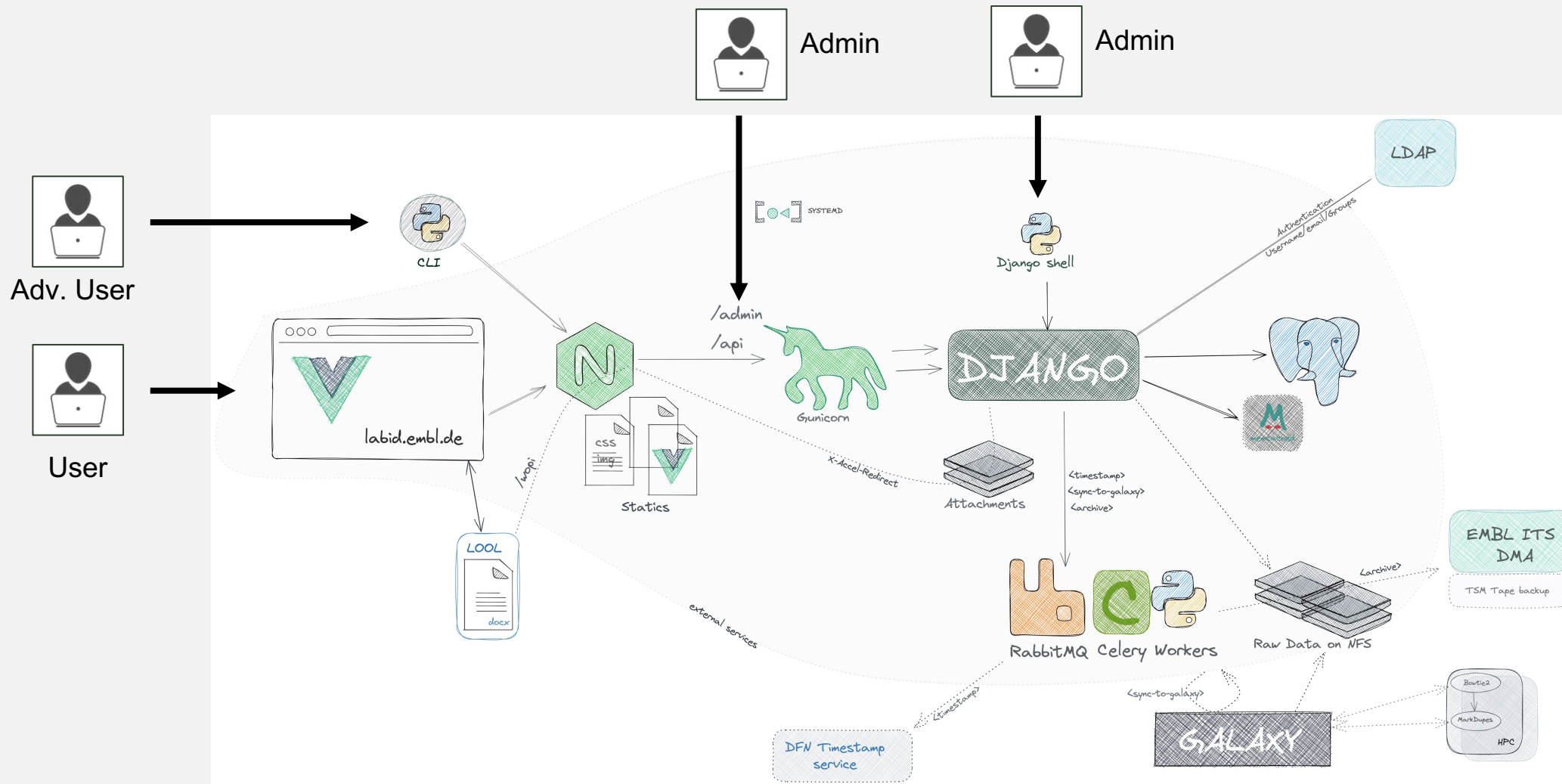
Dataset level

- Standard Metadata extracted (here CZI) “as is” into a read only JSON
- Individual Metadata replicated in editable slots

Assay level

- Metadata common to all datasets are exposed at the assay level

LabID offers several clients and uses a range of services



- The REST API allows to easily develop your own clients

The LabID Command Line Interface

```
(labidcli) (base) mac-furlong98:labid-cli girardot$ labid --help
```

Usage: labid [OPTIONS] COMMAND [ARGS]...

Options

```
--log-level          <debug|info|warning|error|critical>
--log-file           <str>
--modelmeta-cache    <auto|refresh|off>

--modelmeta-cache-expiration <int>
--dma-url             <str>
--dma-scope           <str>
--version             -v
--install-completion
--show-completion
--help
```

Commands

```
assays    Manage assays
config    Create and manage your connection configurations
create    Create core LabID entities (project, study, sample, experiment)
export    Export links and dataset metadata
manage    Manipulate existing items (rename, set-field, annotate, delete)
get       Get items of any types as JSON representation
register   Register raw and derived datasets
validate  Validate 'initialized' assays
job       Manage jobs
transfer  Transfer assay data to LabID library with assay pre-registration
```

- Export Studies to ENA
- Export symbolic links to datasets in your project folder
- Automate raw data registration with cron jobs
- ...

Integrate LabID in your own app with the Python Lib

> Get from <https://gitlab.com/lab-integrated-data/labid-lib>

Usage

After installation, you should be able to use the following imports:

```
import labidlib
from labidlib.core import models
from labidlib.api.manager import LabidManager
```

```
manager: LabidManager = LabidManager(client)
datasets: List[Dataset] = manager.list_datasets(
    filter_type="assay",
    filter_uuid="23bb1a6c-edd1-4478-920b-343c7d34a4f5"
)
if datasets[0].status == DataStatus.ONLINE:
    # do something cool
```

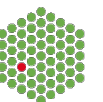
Dozens of methods

```
(m) create_or_update_assay(self, assay)
(m) create_or_update_consumable(self, consumable)
(m) create_or_update_equipment(self, equipment)
(m) create_or_update_experiment(self, experiment)
(m) create_or_update_project(self, project)
(m) create_or_update_sample(self, sample)
(m) create_or_update_specimen(self, specimen)
(m) create_or_update_storage_equipment(self, storage_equipment)
(m) create_or_update_study(self, study)
(m) delete_assay(self, assay)
(m) delete_item(self, item, model, remove_data=False)
(m) delete_items(self, items, model, batch_size=50, remove_data=False, verify=False)
(m) fetch_annotation_id_from_item(self, model, id, ann_type=None)
(m) fetch_annotation_value_from_item(self, model, id, ann_type=None)
(m) fetch_annotationtype_by_name(self, name)
(m) fetch_assay(self, uuid=None, run_dir=None, load_ownership=False)
(m) fetch_cvterm_by_name_and_category(self, name, category_name=None, category_id=None)
(m) fetch_dataset(self, uuid)
(m) fetch_dataset_by_name(self, name, exact_match=True, study_id=None)
(m) fetch_dataset_for_datafile(self, datafile_uuid)
(m) fetch_dataset_for_filename(self, filename, project_uuid=None, study_uuid=None, exact_match=True)
(m) fetch_dataset_for_path(self, path, project_uuid=None, study_uuid=None, exact_match=True, resolve_symlink=True)
(m) fetch_dataset_from_datafile(self, datafile_uuid)
(m) fetch_dataset_id_for_datafile(self, datafile_uuid)
(m) fetch_dataset_id_for_path(self, path, project_uuid=None, study_uuid=None, exact_match=True, resolve_symlink=True)
```

- Use the lib in your own application/script/notebook to register samples, fetch data, launch workflows...
- Use the lib to write your own sniffers



```
@abstractmethod 12 usages (2 dynamic)  👤 charles girardot +1
def sniff_instrument_run_assays(
    self, dir_path: Path, group: str, username: Optional[str] = None
) -> List[InstrumentRun]:
```



Thank you

Multimodal Open Data Integration Support (MODIS) part of EMBL Data-Science Center



Charles Girardot

Head of Multimodal Open Data Integration Support

ORCID: [0000-0003-4301-3920](https://orcid.org/0000-0003-4301-3920)



Laurent Thomas

Postdoctoral Fellow

ORCID: [0000-0001-7686-3249](https://orcid.org/0000-0001-7686-3249)



Md Nayeem Reza

Senior Software Engineer

ORCID: [0000-0003-2068-5812](https://orcid.org/0000-0003-2068-5812)



Matthias Monfort

Web Developer / Workflow Management

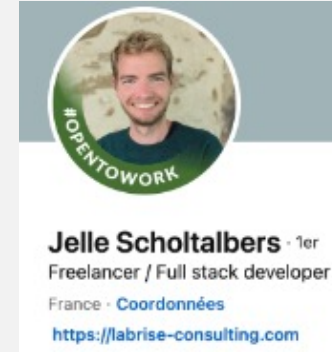


Marco Wetter

Full Stack Developer

Contact : modis@embl.de

Chat: <https://chat.embl.org/embl/channels/labid>



Join LabID on Slack