

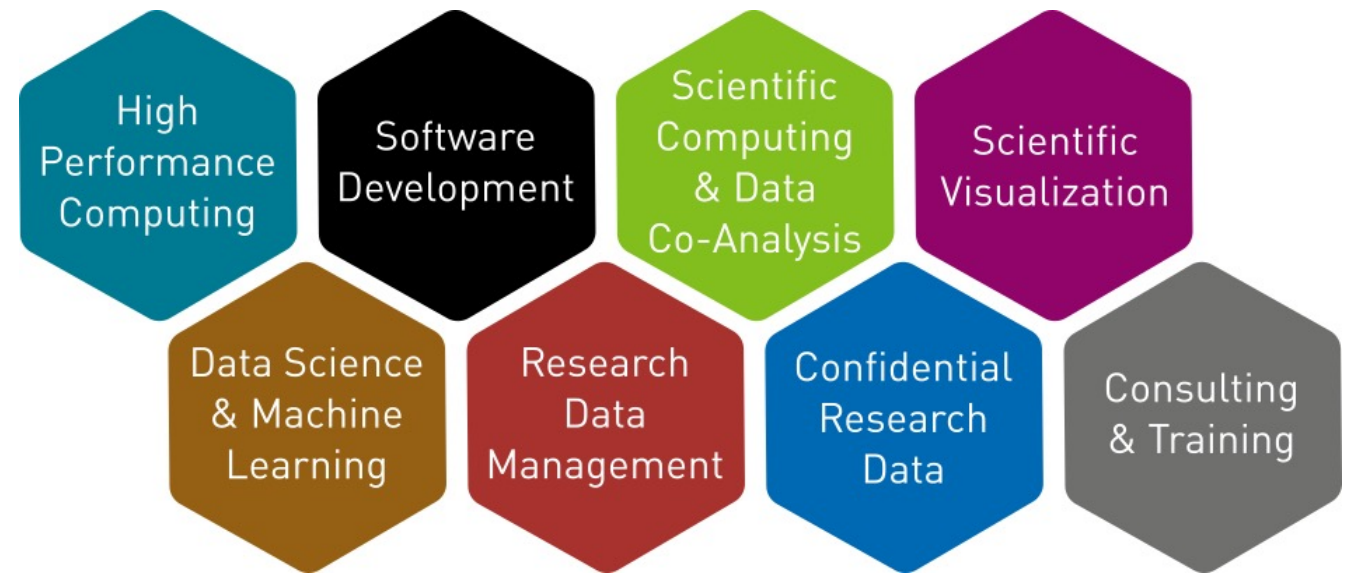


openRDM.eu: FAIR and reproducible
data management and analysis with
openBIS

Caterina Barillari
Scientific IT Services, ETH Zurich



Scientific IT Services of ETHZ



- ❑ A section of ETHZ IT Services
- ❑ About 40 experts in various areas of scientific computing
- ❑ With a background in different areas of science

A European RDM service for the academic community

- ❑ Project funded by EGI-ACE
- ❑ Timeline: January 2021- December 2022



Self-hosted openBIS RDM platform
Support for set up on local IT infrastructure

RDM= Research Data Management



Training & 'best effort' user support

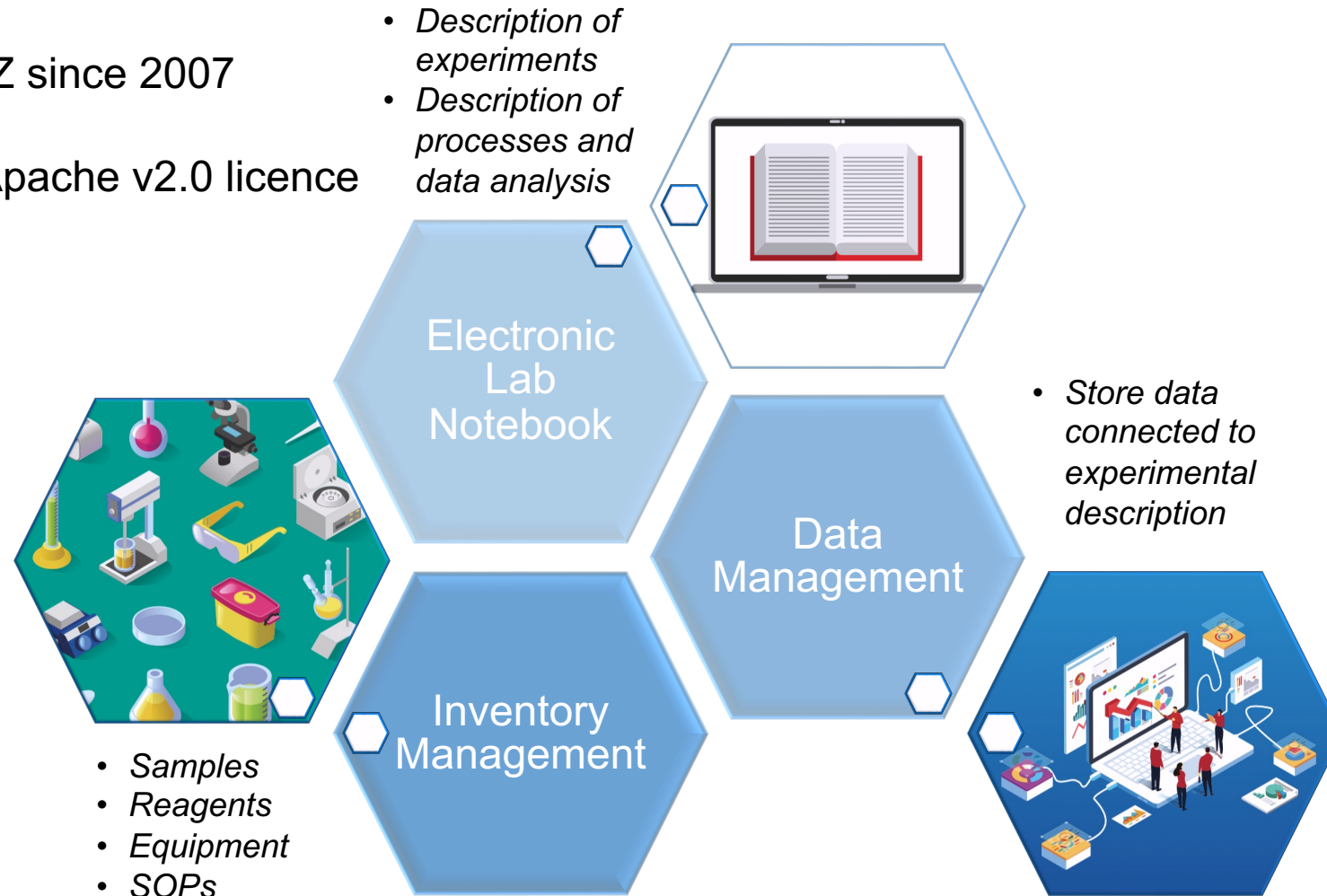


HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

openRDM.eu customers

openBIS: a complete solution towards FAIR data management

- ❑ Developed at ETHZ since 2007
- ❑ Distributed under Apache v2.0 licence



openBIS Inventory management



Lab equipment

Lab samples & materials

Collection: Mixers

Code	Name	Identifier	Mixing volume	Log Number	Year of registration	Notes	Type	Space	Parents
EQUIPMENT_MIXERS_6	Engh Pul Plus	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_6	10 L	LOG 139-21.006	2019		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_7	Hobart 50	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_7	8 L	LOG 139-22.901			Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_1	Tecator inclusion vacuum mixer	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_1		LOG 139-29.009	2008		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_2	Tecator inclusion vacuum mixer	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_2		LOG 308-29.007	2016		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_3	ENVA R 28 M	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_3	75 L	LOG 139-21.003	1991		Equipment Mixers	EQUIPMENT	EQUIPMENT/EQUIPMENT_MIXERS_3 (Change of st - example)
EQUIPMENT_MIXERS_4	ENVA RV 11	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_4	390 L	LOG 139-21.004	1999		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_5	Engh R 28 M Spix	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_5	75 L	LOG 139-21.005	2012		Equipment Mixers	EQUIPMENT	

Collection: Chemical admixtures

Code	Name	Identifier	Admixture type	Admixture type other	Date of reception	Manufacturer name	Production date	Mass volumetric density	Solids content	Notes	Type	Space
ADM6	Skagart 705 L	MATERIALSRAW-MATERIALSADM6		OTHER	2021-06-25 14:20:09 +0000	Sika		800.0			Admixture	WATERU
ADM7	Si-O STARVIS S 2100 F SAP	MATERIALSRAW-MATERIALSADM7		OTHER	2020-04-20 21:00:00 +0000	SASF					Admixture	WATERU
ADM8	Ono 400	MATERIALSRAW-MATERIALSADM8	Retarder		2017-09-27 07:00:00 +0000						Admixture	WATERU
ADM9	Sika Wandfest superplasticizer	MATERIALSRAW-MATERIALSADM9	Superplasticizer		2019-12-19 15:48:33 +0000	Sika	2020-01-02 15:48:33 +0000	1080.0			Admixture	WATERU
ADM10	Lithium Carbonate	MATERIALSRAW-MATERIALSADM10	Accelerator								Admixture	WATERU
ADM11	No W.A. 130	MATERIALSRAW-MATERIALSADM11	Superplasticizer								Admixture	WATERU

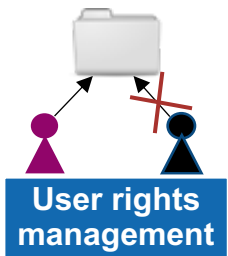
Lab procedures



Samples' storage manager



Barcode reader



User rights management

Collection: Shrinkage

Code	Name	Identifier	Shrinkage dimensionality	Type of time scale (if measured values)	SOP ID	Notes	Type	Space
SHRINKAGE_PROTOCOL_1	SA 3521, Appendix F	METHODSMEASUREMENT_PROTOCOLS-SHRINKAGE_PROTOCOL_1	Linear shrinkage	Unstructured grid (varying time steps) linear scale	4003	Protocol for measuring linear shrinkage of concrete specimens according to the Swiss standard SA 3521 (Appendix F).	Shrinkage Protocol	METHODS

openBIS Electronic Lab Notebook

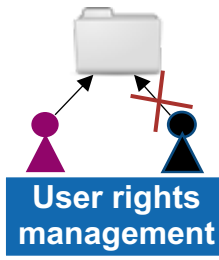
Experimental Step: Detection of LexA-ER-B42 induction by flow cytometry

Name: Detection of LexA-ER-B42 induction by flow cytometry
 Owner: Diana Ottoz
 Experimental goals: Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout

Parents

Name	Identifier	Comments	Organism	Storage conditions	Stock concentration	Sterilization	Publication	Protocol type	Materials	Time requirement
LexA-ER-B42 + target	/MATERIALS/YEASTS/FRY418						Ottoz et al., Nucleic Acids Research, 2014			
1000X Cycloheximide	/MATERIALS/REAGENTS/FRSOB34	diluted 1/1000, treated for 1/2 hour		+4 degrees	1000X	none				

Personal folder



Entities relations

Audit trail

Entity history

History of Yeast: FRY1

1-10 of 10 rows per page: 10 COLUMNS FILTERS EXPORTS

Version	Author	Changes	Full Document	Timestamp
1	admin	Common markers [COMMON_MARKERS] property changed: URA3_HIS3_LEU2 [COMMON_MARKERS]	show	2015-01-26 16:55:03
2	admin	Background-specific markers [BACKGROUND-SPECIFIC_MARKERS] property changed: MET15 [BACKGROUND_SPECIFIC_MARKERS] Endogenous 2micron plasmid in yeast [ENDOGENOUS_PLASMID] property changed: CIR_PLUS [ENDOGENOUS_PLASMID] Yeast strain name [YEAST_STRAIN_NAME] property changed: BY4741	show	2015-01-26 16:55:03
3	admin	Source [SOURCE] property changed: Euroscarf	show	2015-01-26 16:55:03
4	admin	Mating Type [MATING_TYPE] property changed: A [MATING_TYPE] Owner [OWNER] property changed: FABIAN_RUDOLF [OWNER]	show	2015-01-26 16:55:03
5	admin	Annotations State [ANNOTATIONS_STATE] property changed: <root> Genetic Background [GENETIC_BACKGROUND] property changed: BY4741 [GENETIC_BACKGROUND]	show	2015-01-26 16:55:03
6	diana	CHILD relation(s) added: 20150126165503379-125	show	2015-02-20 14:21:00

Deletion history

TOOLS

History DELETION X

Dynamic Property Plugins
Entity Validation Plugins
Queries
History

Entity Type	Entity Identifier	Entity Space	Entity Project	Entity Registrar	Entity Registration Date	Reason	Description	Contact	User	Date
DELETION	2021092112194905-49	DEFAULT_LAB_NOTEBOOK	DEFAULT_LAB_NOTEBOOK.DEFAULT_PROJECT	admin	2021-09-21 14:19:50	test for deletion history	2021092112194905-49	show	admin	2021-09-21 17:54:42
FREEZING	20210921121917398-48	MATERIALS	MATERIALS.BACTERIA	admin	2021-09-21 14:19:17	duplicate	20210921121917398-48	show	admin	2021-09-21 17:52:17

Rows per page: 10 1 of 2

Data ingestion into openBIS

Web UI (low GBs)

Python API (low GBs)

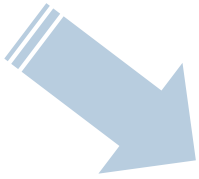
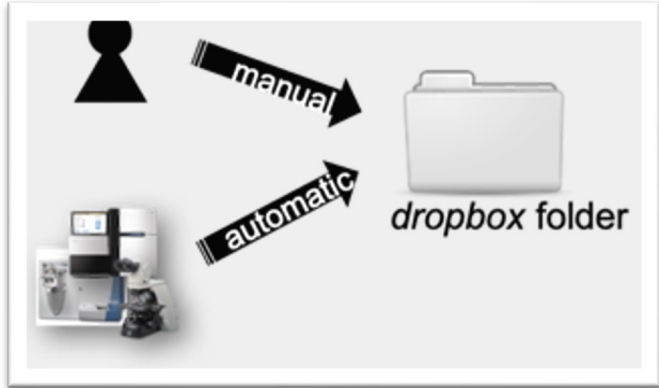
“Dropbox” (high GBs)

Experimental Step: Detection of LexA-ER-B42 induction by flow cytometry

General

Experimental results:

```
ds_new = o.new_dataset(  
    type = 'ANALYZED_DATA',  
    experiment = '/SPACE/PROJECT/EXP1',  
    sample = '/SPACE/SAMP1',  
    files = ['my_analyzed_data.dat'],  
    props = {'name': 'some good name', 'description': '...' }  
)  
ds_new.save()
```



Data

Experimental Step: Detection of LexA-ER-B42 induction by flow cytometry

Experimental results:

Identifier	Comments	Details	Suggestions	Type	Modification Date
ANALYZED_DATA/STEP/...	LexA-ER-B42 - target			Node	2022-05-10 16:41:28
ANALYZED_DATA/STEP/...	about 11000, treated for 12 hour	Disable in DMSO, Repeat		Buffer	2022-05-25 17:07:27
ANALYZED_DATA/STEP/...	1:2 concentration series with 2000 uM	Disable in		Solution	2022-05-25

Data analysis: openBIS & Jupyter Notebooks

Space: Barillac Barillac

- + New Project
- More ...
 - Export Metadata
 - Export Metadata & Data
 - New Jupyter notebook

```
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 C
```

Datasets Download

```
In [12]: ds0.download(files=ds0.file_list, destination='./', wait_until_finished=True)
Files downloaded to: ./20190328133056219-504
Out[12]: './'
```

Read and display the data

```
In [ ]: img = imread('./20190328133056219-504/original/blobs.tif')
fig = plt.figure(figsize=(10,5))
fig.add_subplot(121)
plt.imshow(img, cmap='gray')
plt.title('blobs')
fig.add_subplot(122)
plt.hist(img.ravel(), bins=10)
plt.title('Histogram of blobs')
plt.show()
```

Saving the results

The next cell stores a copy of this notebook as an html file so it can be open easily in browsers without the need of Jupyter, save before executing this to get the latest version as html

```
In [ ]: from nbconvert import HTMLExporter
import codecs
import nbformat
exporter = HTMLExporter()
output_notebook = nbformat.read(fileName, as_version=4)
output_resources = exporter.from_notebook_node(output_notebook)
codecs.open(fileName + '.html', 'w', encoding='utf-8').write(output)
```

Sets the owner of the result dataset

```
In [ ]: owner= o.get_object('/SIS_HLUETCKE/DEMO/JUPYTER-DEMO'),
owner
```

Creates the result dataset

openBIS + JupyterHub server

Experimental Step: jupyter-demo

- + New
- Edit
- Upload
- More ...

General info

Name: jupyter-demo

Show in project overview: true

Start date: 2020-01-14 11:27:08 +0000

- RAW_DATA : 20200114134628751-71
 - blobs.tif (63.6 kb)
- ANALYZED_DATA : jupyter notebook demc
 - test 5.ipynb (6.6 kb)
 - test 5.ipynb.html (281.8 kb)

JupyterLab-openBIS extension:
<https://pypi.org/project/jupyterlab-openbis/>

Data publication: export to Zenodo

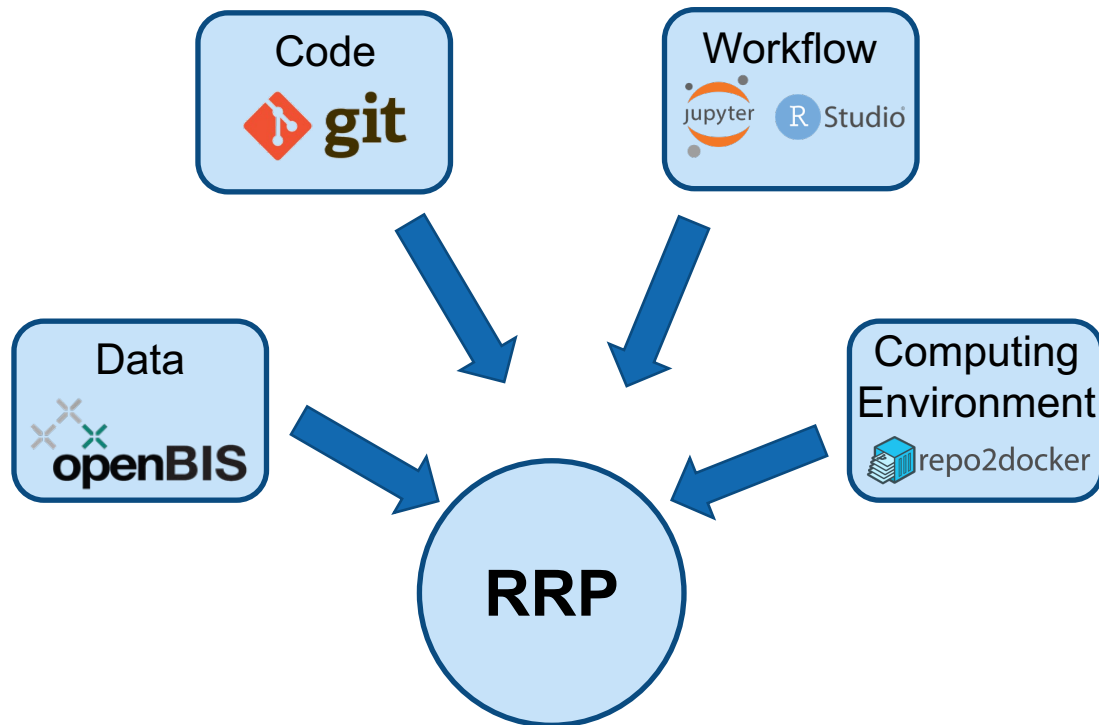
The screenshot shows the 'Zenodo Export Builder' interface. On the left is a navigation menu with categories like Lab Notebook, Inventory, Stock, Utilities, Exports, Storage Manager, User Manager, Trashcan, Settings, Other Tools, and About. The 'Exports' section is expanded, showing 'Export to Zenodo' as the selected option. The main area displays a tree view of the openBIS structure. A blue arrow points from the 'Inducible Transcription Factor' folder to the 'Export Selected' button. Below the tree view, there are instructions: 'You can select any parts of the accessible openBIS structure to export: If you select a tree node and do not expand it, everything below this node will be exported. To export selectively only parts of a tree, open the nodes and select what to export.' and 'Publication time constraint: After the resource has been exported it should be published in Zenodo UI within 2 hours. Otherwise, the publication metadata will not be registered in openBIS.'

On the right, the Zenodo 'New upload' form is visible. It includes a search bar, a 'Save' button, and a 'Publish' button. The form has several sections: 'Files' (showing a file named 'content.zip' of size 1.4 MB), 'Communities' (with a search input), 'Upload type' (with options like Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, Physical object, Workflow, Other), and 'Basic information' (with fields for Digital Object Identifier, Publication date, and Title). A blue box with a white dot and the text 'Title' points to the 'Title' field in the 'Basic information' section.

☐ Integration also with ETH data repo

Reproducible Research Platform (RRP)

Computational reproducibility with established open source tools:



- ❑ openBIS for **research data** management
- ❑ Git for **code** management
- ❑ JupyterLab (RStudio) for documenting the **workflow**
- ❑ repo2docker for the **computing environment**
→ easy to build container images (conda, R, Python, Julia)
- ❑ RRP allows collaboration
- ❑ Elastic compute backend (Kubernetes)

openBIS demo

Example use-case: RNA sequencing study of 8 different dog breeds

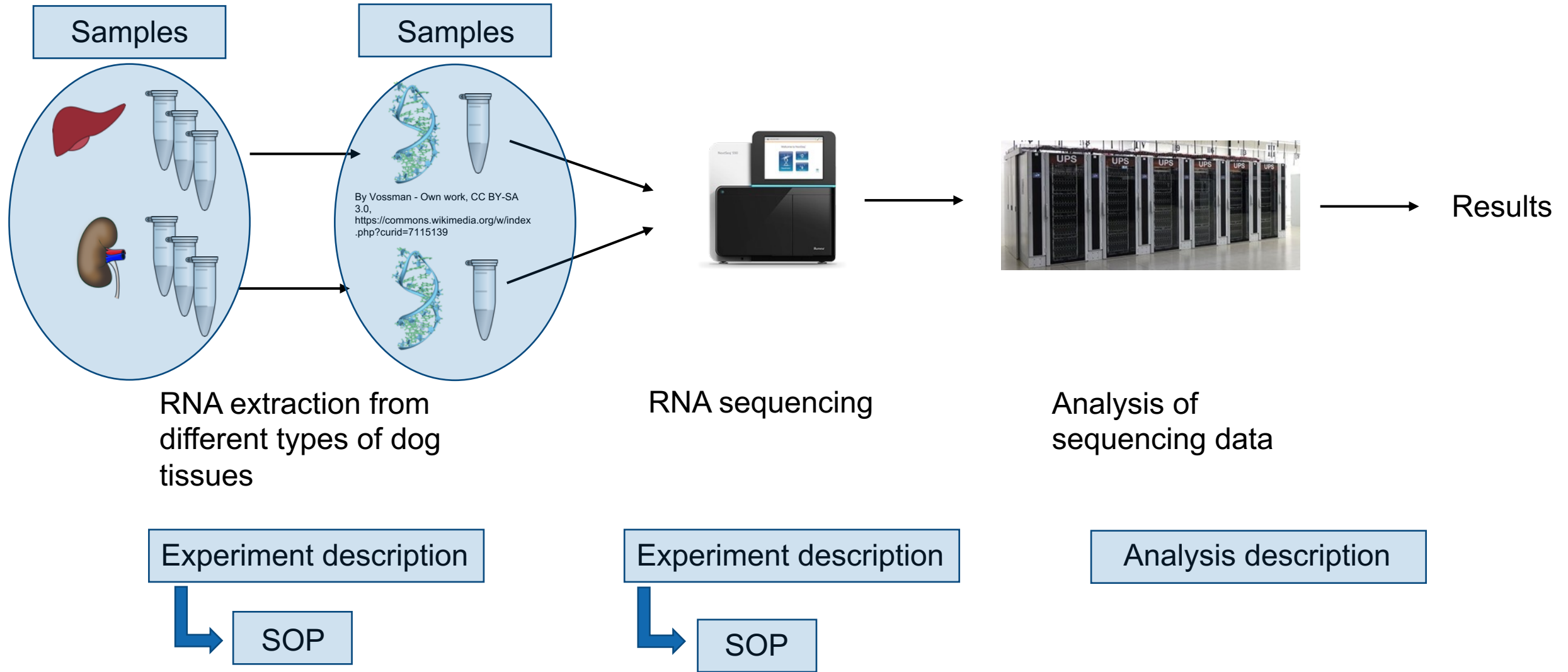
Goal of the study: *understand intra breed genetic variability of dogs with special reference to Beagle dog, since Beagles are used as an animal model for compound testing in the pharma industry.*

The eight different dog breeds are as follows:










1. Beagle
2. GSD
3. Golden Retriever
4. Terrier
5. King Charles
6. Poodle
7. Rottweiler
8. West Highland White Terrier



Overview of the study process












How does the process look like in openBIS?

-  **Inventory**
 -  Barillac Materials
 -  Samples
 -  Chemicals
 -  RNA extracts
 -  Tissues
 -  Barillac Methods
 -  Protocols
 -  RNA extraction protocols

Samples

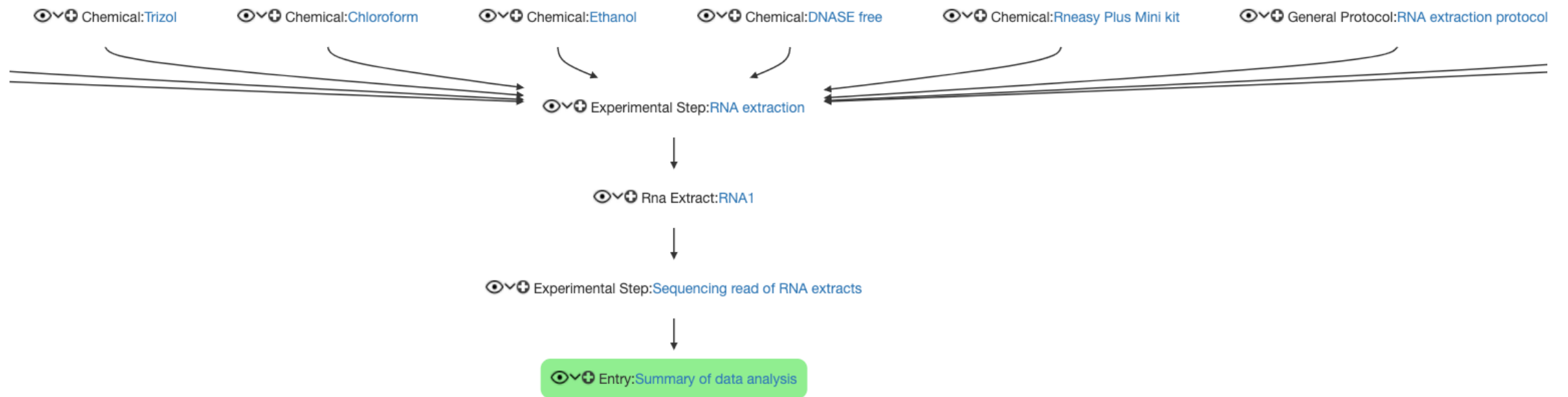
SOP

-  Lab Notebook
 -  My Space (Barillac Barillac)
 -  Rna Sequencing Of 8 Dog Breeds
 -  RNA sequencing
 -  RNA extraction
 -  Sequencing read of RNA extracts
 -  RNA seq data
 -   Summary of data analysis

Experiment description

Analysis description

How does the process look like in openBIS?



Acknowledgements – openBIS & RDM team



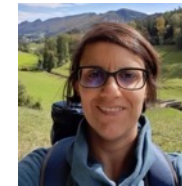
Juan Fuentes



Piotr Kupczyk



Viktor Kovtun



Caterina Barillari



Henry Lütcke



Swen Vermeul



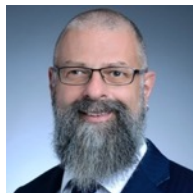
Richard Wartenburger



Artur Pedziwilk



Bernd Rinn
SIS Section Head



Matthew Baker
SSDM Group Head



Sergio Maffioletti
RP Group Head



Thomas Wüst
CDSS Group Head

Contacts & useful info

Documentation & video tutorials: <https://openbis.ch>

SIS website: <https://sis.id.ethz.ch/>

Twitter: https://twitter.com/ETH_SIS

openBIS support

openbis-support@id.ethz.ch