



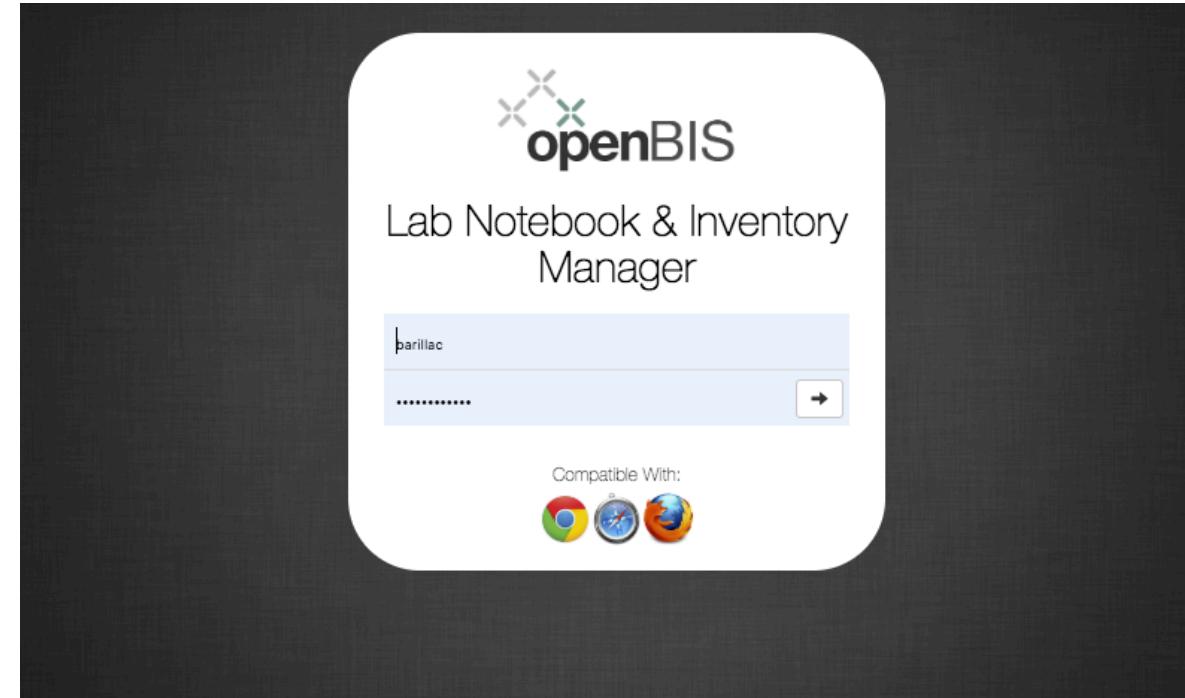
Introduction to Research Data Management with openBIS

Caterina Barillari & Priyasma Bhoumik

Scientific IT Services, ETH Zurich

ETH Zurich, 17.11.2020

<https://openbis-training.ethz.ch/openbis/webapp/eln-lims/?>



Overview of training

1. Introduction to basic openBIS concepts
2. Hands-on tutorial:
 1. Registration of samples in openBIS Materials Inventory
 2. Registration of protocols in openBIS Methods Inventory
 3. Recording experiments and uploading data in openBIS Lab notebook
 4. Searching the ELN
 5. Freezing entities
 6. Jupyter notebooks

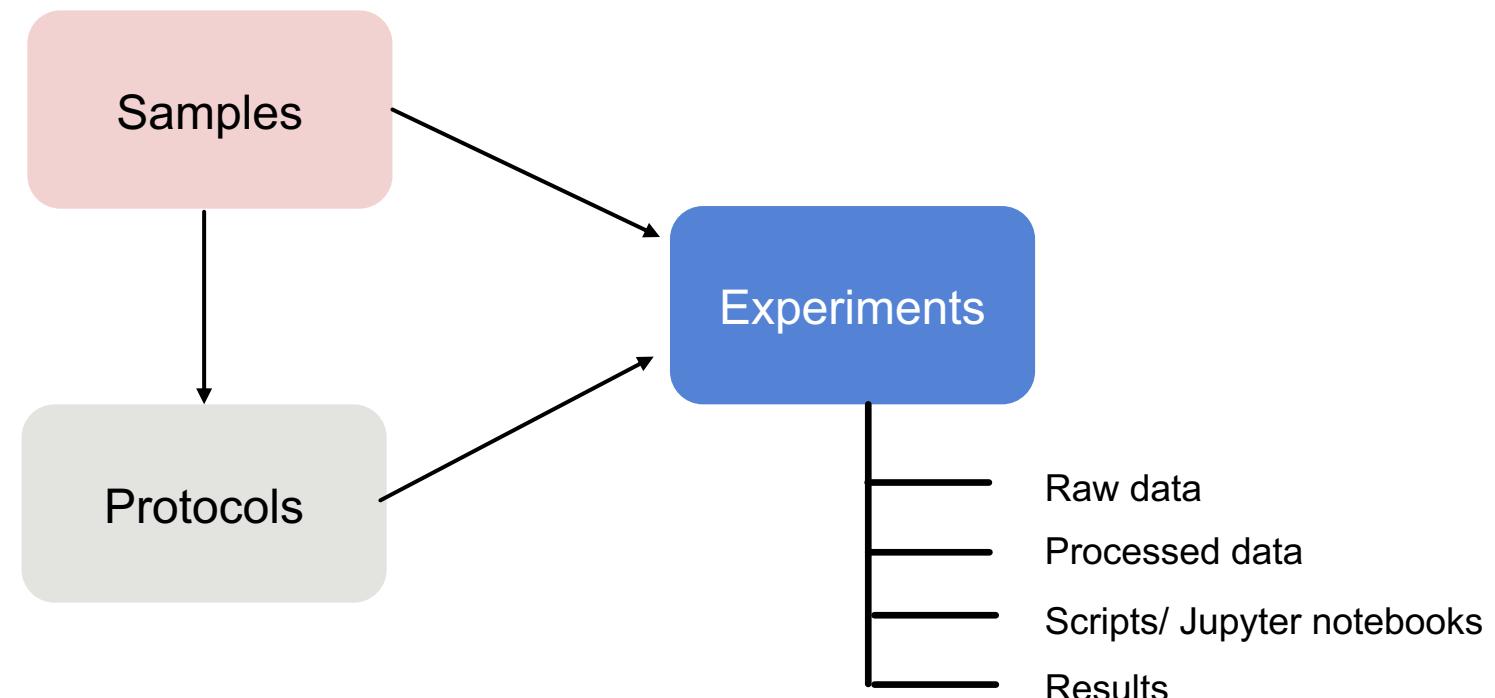
How to use the openBIS Inventory & Lab Notebook

Inventory

Shared by all lab members.

Lab Notebook

Personal space. Can be shared with colleagues/collaborators.



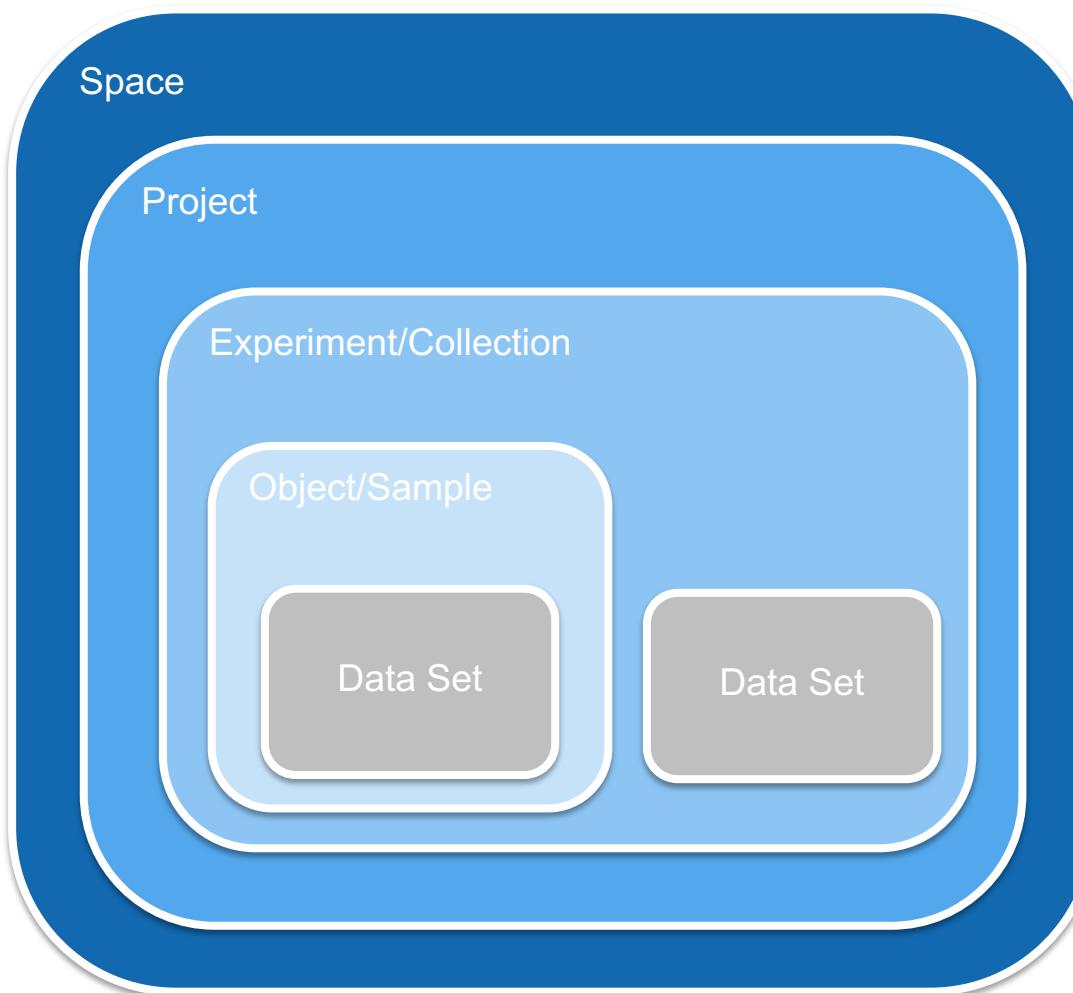
openBIS Inventory

- The default installation has two main folders in the Inventory:
 - **Materials:** all samples and materials can be stored in collections in this folder
 - **Methods:** all lab protocols (if used!) can be stored in collections in this folder

Organization of openBIS Lab Notebook

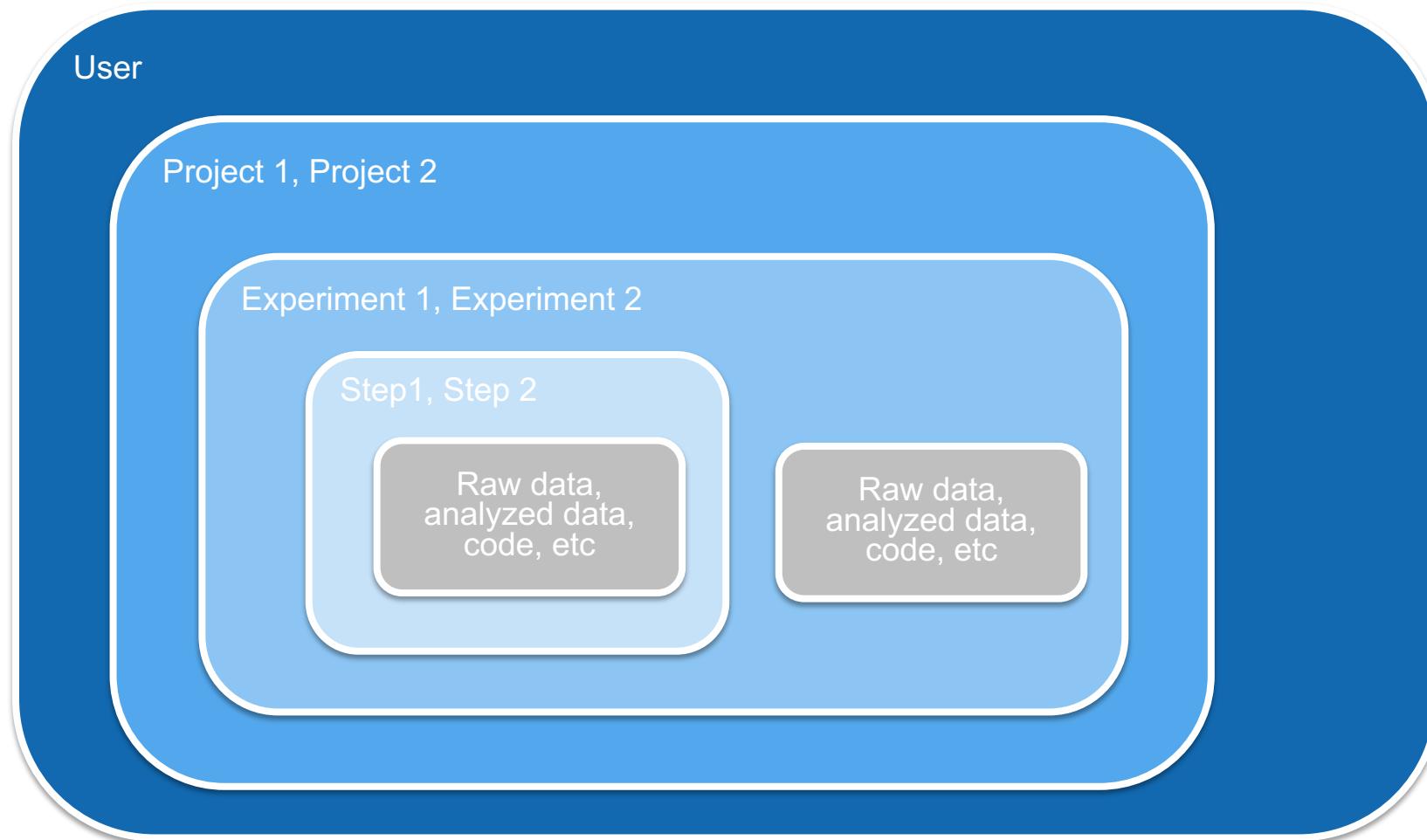
- In the Lab Notebook part of openBIS, usually each user has a **personal Space** where to organize **Projects** and **Experiments**
- An openBIS **Experiment** is a specific scientific question. The single attempt to answer this question can be modelled as **Experimental Steps**.
- **Experimental Steps** can be linked to Samples, Protocols, other Experimental Steps
- Data (raw, processed, analysed, final results) can be attached to Experiments or Experimental steps in **Datasets**

openBIS data structure



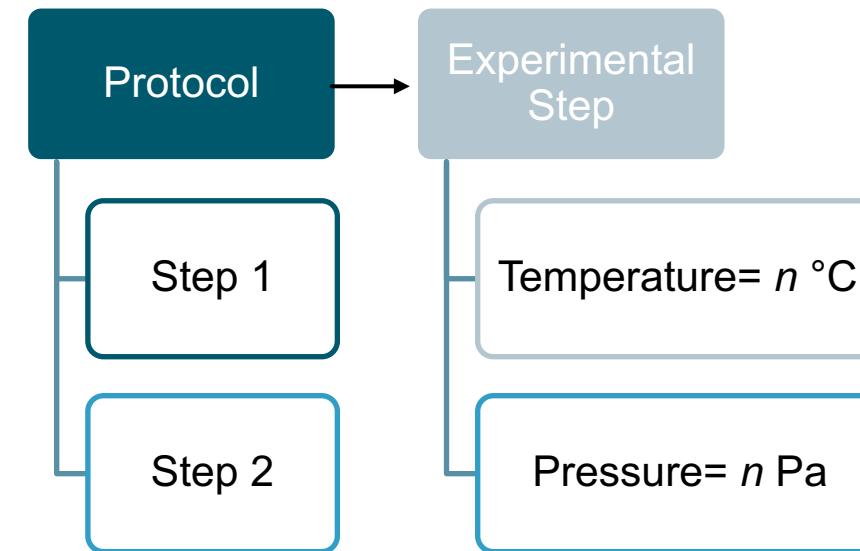
1. **Folder with description**
2. **Folder with description**
3. **Folder with user-defined properties.** There can be several types of Experiment/Collection, each defined by different properties. Example: *Microscopy experiment, PCR experiment etc.*
4. **Entity with user-defined properties.** There can be several types of Objects/Samples, each defined by different properties. Examples: *Antibody, Chemical, Sensor, Chip, General Protocol, Experimental Step...*
5. **Folder for storing data files with user-defined properties.** There can be several types of Data Sets, each defined by different properties.

openBIS Lab notebook



Protocols or Experimental Steps?

- Protocols are standard procedures used in the lab that need to be shared with all lab members
- Every time a given protocol is followed when performing one Experimental Step, the protocol can be linked as parent and the experimental details should be recorded in the Experimental Step itself.
- If you do not follow standard procedures, you do not need to use the Inventory of Methods



Example

Protocol: muffins recipe

Ingredients	Method
2 medium eggs	1. Heat oven to 200C/180C fan/gas 6. Line 2 muffin trays with paper muffin cases. In a large bowl beat 2 medium eggs lightly with a handheld electric mixer for 1 min.
125ml vegetable oil	
250ml semi-skimmed milk	
250g golden caster sugar	2. Add 125ml vegetable oil and 250ml semi-skimmed milk and beat until just combined then add 250g golden caster sugar and whisk until you have a smooth batter.
400g self-raising flour (or same quantity plain flour and 3 tsp baking powder)	
1 tsp salt	3. Sift in 400g self-raising flour and 1 tsp salt (or 400g plain flour and 3 tsp baking powder if using) then mix until just smooth. Be careful not to over-mix the batter as this will make the muffins tough.
100g chocolate chips or dried fruit such as sultanas or dried cherries (optional)	4. Stir in 100g chocolate chips or dried fruit if using.
	5. Fill muffin cases two-thirds full and bake for 20-25 mins, until risen, firm to the touch and a skewer inserted in the middle comes out clean. If the trays will not fit on 1 shelf, swap the shelves around after 15 mins of cooking.
	6. Leave the muffins in the tin to cool for a few mins and transfer to a wire rack to cool completely.

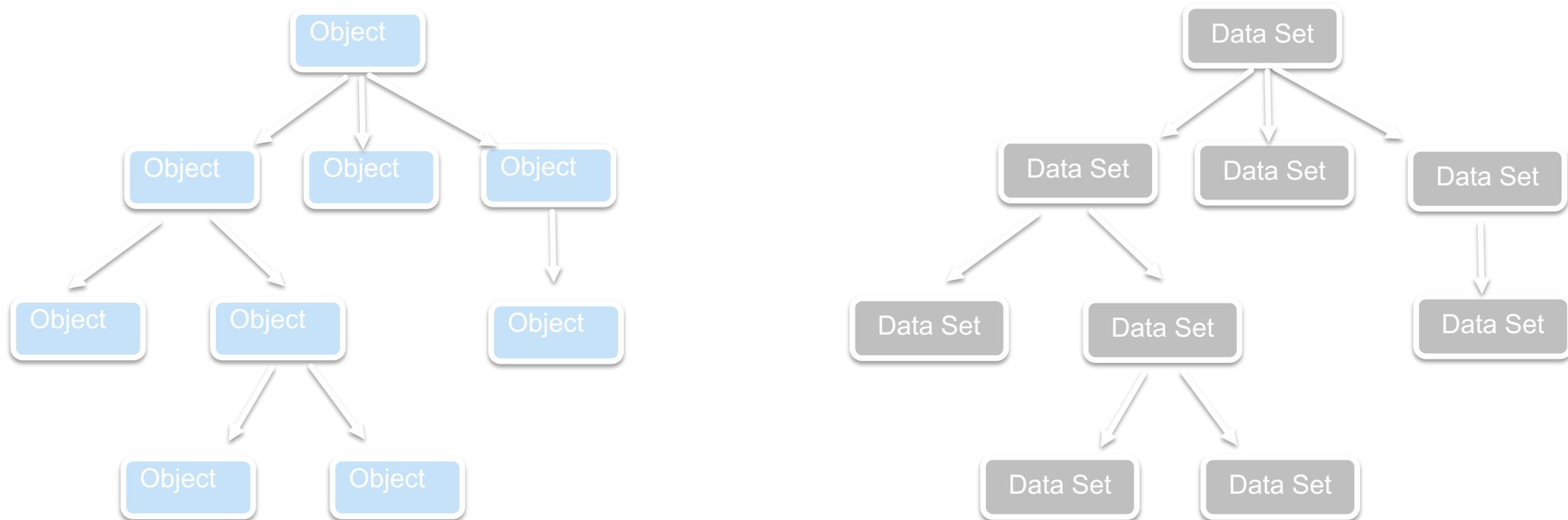
Experimental Step: making muffins

- I use 240 ml of whole milk, instead of 250 ml semi-skimmed
- I add a teaspoon of sodium bicarbonate
- I cook the muffins for 30 mins at 180°C

The protocol gives me directions and defines the steps; When I perform my experiment I can change some parameters and this needs to be recorded in the experimental description.

Linking objects and datasets

- openBIS objects can be linked to other objects and datasets to other datasets with $N:N$ relationship
- In openBIS terms, these are “parent-child” relationships

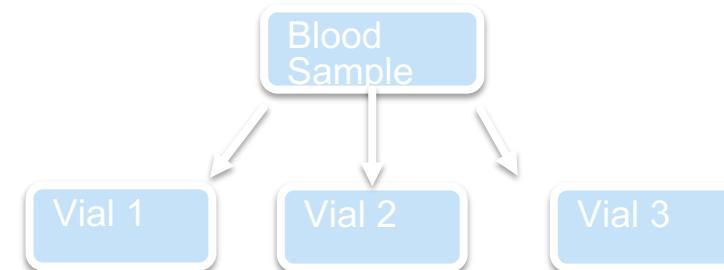


What are “parents” and “children”?

- They are a way of connecting entities together.

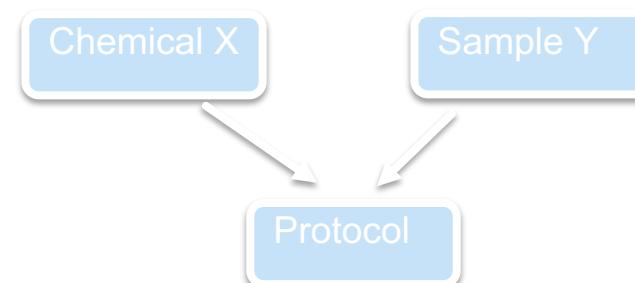
- Examples:

1. One sample is split into several vials and each of them is used for different types of measurements.



The Blood Sample is parent of the samples in each vial

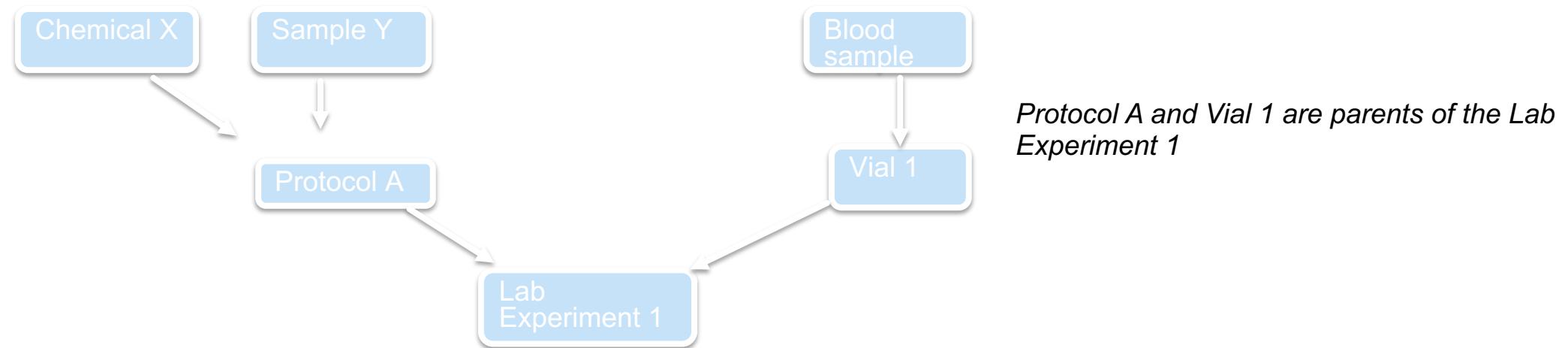
2. You write a protocol, and want to keep track of the samples used.



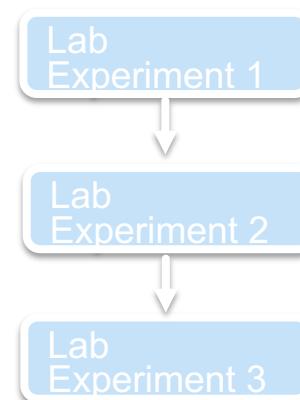
Chemical X and Sample Y are parents of the protocol

What are “parents” and “children”?

3. You describe an experiment and you want to keep track of the protocol(s) and samples used



4. You make one experiment in several steps and you want to link each step to the previous one



Roles

Roles	Capabilities
Observer (Space or Instance)	Read-only access to given spaces or to the whole instance.
Space/Project user	Create + edit Object, Experiment. Edit Project .
Space/Project power user	User rights +create Projects . Delete object, experiment, project, datasets. Add, update vocabulary terms.
Space/Project admin	Power user rights+ list roles; create and delete space roles; edit datasets.
Instance admin	Space admin rights + create types. Has access to everything.

Freezing entities

It is possible to “freeze” every level of the openBIS hierarchy.

Space: Default Lab Notebook

+ New Project More ... ▾

- Export Metadata
- Export Metadata & Data
- New Jupyter notebook
- Manage access
- Freeze Entity (Disable further modifications)

Freezing entities

At every level, everything below is always selected to be frozen. Selection can be modified.

Freeze Entity

Choose the entities to freeze (all by default):

Selected	Type	PermId	Name
<input checked="" type="checkbox"/>	Space	DEFAULT_LAB_NOTEBOOK	DEFAULT_LAB_NOTEBOOK
<input checked="" type="checkbox"/>	Project	20190528221459905-1	DEFAULT_PROJECT
<input checked="" type="checkbox"/>	Project	20190611122038970-49	DEMO_PROJECT
<input checked="" type="checkbox"/>	Experiment/Collection	20190611122056078-50	Demo experiment 1
<input checked="" type="checkbox"/>	Experiment/Collection	20190528221459905-14	Default Experiment
<input checked="" type="checkbox"/>	Object	20190611122145679-52	Step 2
<input checked="" type="checkbox"/>	Object	20190611122213447-53	Step 3
<input checked="" type="checkbox"/>	Object	20190611122127059-51	Step 1
<input checked="" type="checkbox"/>	Object	20190528221828470-42	test
<input checked="" type="checkbox"/>	DataSet	20190611123601400-54	test data

Enter your password to freeze the entities, after they are frozen no more changes will be allowed:
This operation is irreversible!

Password (*):

Accept Cancel



Barcode Generator

- Barcodes can be generated & printed after enabling this feature from Settings

The screenshot shows the Barcode Generator interface. On the left is a sidebar with navigation links: Lab Notebook, Inventory, Stock, Utilities (with sub-links like Jupyter Workspace, New Jupyter Notebook, User Profile, Barcodes Generator, Object Browser, Vocabulary Browser, Advanced Search, Exports, Storage Manager, User Manager, Trashcan, Settings, and About). The 'Barcodes Generator' link is highlighted with a blue box. The main area has five buttons at the top: a power button, a search bar labeled 'Global S...', and three download-related icons. Below this is a form with fields: 'Type:' set to 'Code 128', 'Count:' set to '10', 'Layout:' set to 'Split Layout', 'Width:' set to '50 mm', 'Height:' set to '15 mm', a 'Generate Barcodes' button, and a printer icon. Underneath the form are four generated barcode images, each with a unique identifier below it: 20200916100807191-692, 20200916100807191-693, 20200916100807191-694, and 20200916100807191-695.

- It is possible to assign existing barcodes to samples

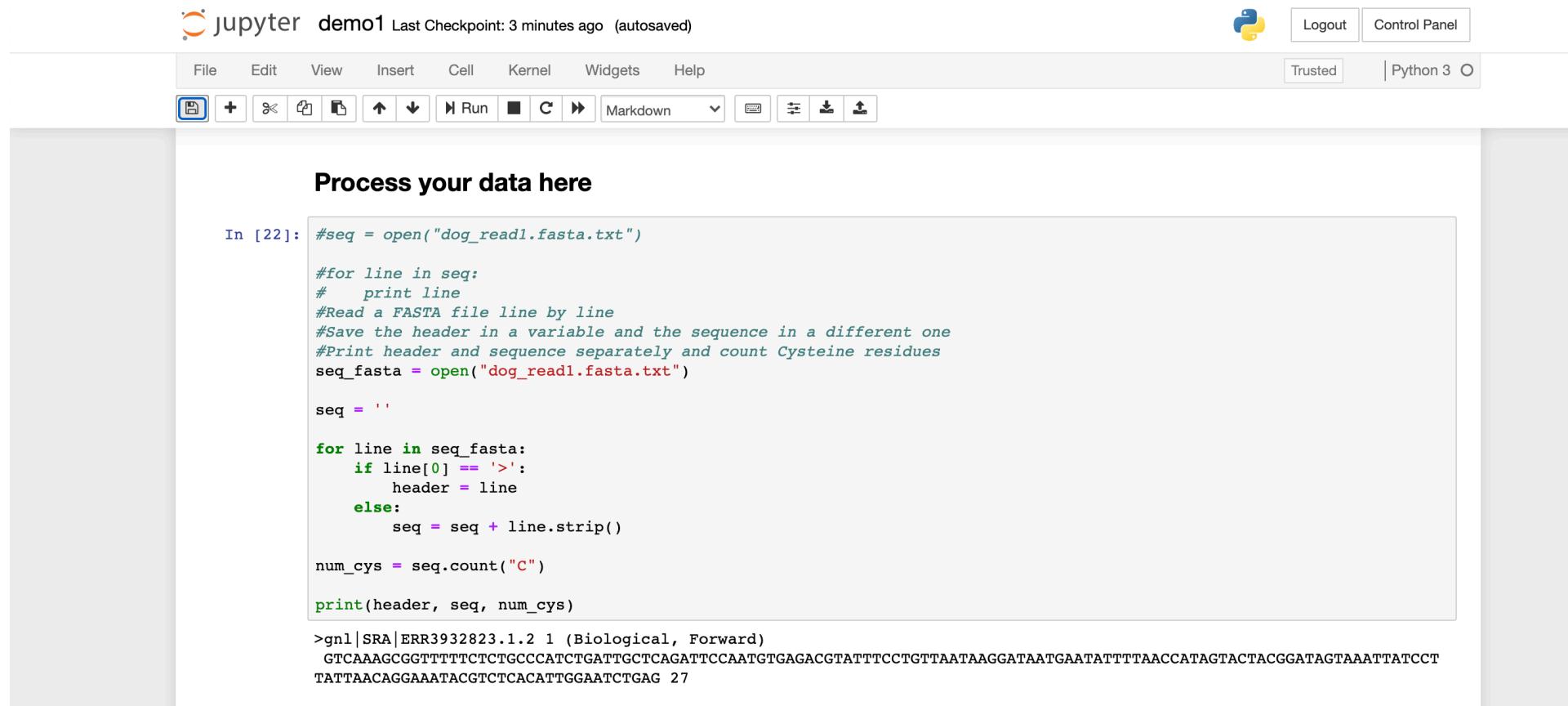
Jupyter Notebooks

- Jupyter notebooks can be used locally or via JupyterHub server for downstream data analysis.

The screenshot shows the Eln-Lims interface for managing experiments. The main title is "Default Experiment: RNA analysis of 8 dog breeds". On the left, there's a sidebar with various lab management links like Lab Notebook, My Space, Intra Breed Genetic Variability Of Dogs, and Utilities. The "Intra Breed Genetic Variability Of Dogs" section is expanded, showing a folder named "RNA analysis of 8 dog breeds" which is selected. A context menu is open over this folder, with the option "New Jupyter notebook" highlighted with a red box. The main panel displays experiment details: Name (RNA analysis of 8 dog breeds), Show in project overview (true), Start date (2020-06-24 20:49:21 +0000), Experimental details (Goals: Understand intra breed genetic variability of dogs with special reference to Beagle dog), and References (Publication: Lindblad-Toh K et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819).

Jupyter Notebooks

- Jupyter notebooks can be used locally or via JupyterHub server for downstream data analysis.



The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** jupyter demo1 Last Checkpoint: 3 minutes ago (autosaved)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Run, Markdown, Trusted, Python 3
- Section Header:** Process your data here
- Code Cell (In [22]):**

```
#seq = open("dog_read1.fasta.txt")
#Read a FASTA file line by line
#Save the header in a variable and the sequence in a different one
#Print header and sequence separately and count Cysteine residues
seq_fasta = open("dog_read1.fasta.txt")

seq = ''

for line in seq_fasta:
    if line[0] == '>':
        header = line
    else:
        seq = seq + line.strip()

num_cys = seq.count("C")

print(header, seq, num_cys)
```
- Output:**

```
>gnl|SRA|ERR3932823.1.2 1 (Biological, Forward)
GTCAAAGCGGTTTCTGCCATCTGATTGCTCAGATTCCAATGTGAGACGTATTCCTGTTAATAAGGATAATGAATATTTAACCATAGTACTACGGATAGTAAATTATCCT
TATTAACAGGAAATACGTCACATTGGAATCTGAG 27
```

Data ingestion into openBIS

1. Via web interface

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

General

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

Experimental results:
The LexA-ER-B42 induction can be measured by using a target gene encoding a fluorescence protein. LexA-ER-B42 induction is different at different concentrations of inducer.



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

Experimental goals:
Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout

Experimental results:
The LexA-ER-B42 induction can be measured by using a target gene encoding a fluorescence protein. LexA-ER-B42 induction is different at different concentrations of inducer.

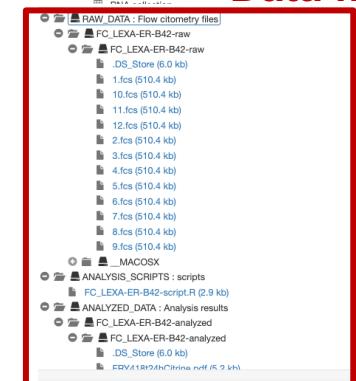
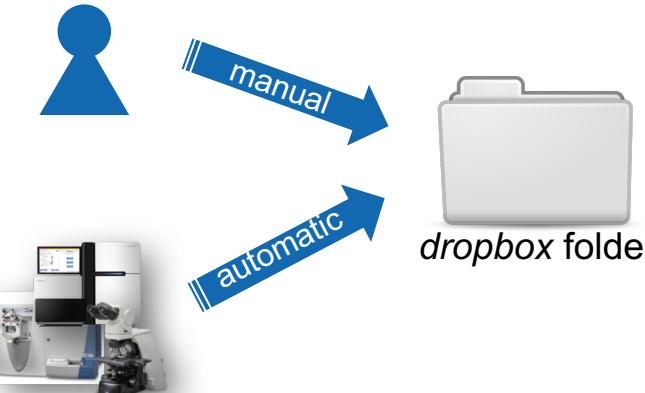
FRY418 24 h induction

Data files

Parents

Identifier	Comments	Details	Suggestions	Type	Modification Date
/MATERIALS/YEASTS/FRY418	LexA-ER-B42 + target			Yeast	2020-04-30 16:41:38
/MATERIALS/REAGENTS/FRSOB34	diluted 1/1000, treated for 1/2 hour	Dissolve in DMSO. Aliquot.		Solution Buffer	2017-11-20 17:07:37
/MATERIALS/REAGENTS/FRSOB37	1:2 concentration series with 2000 nM maximum for 24 hours	Dissolve in EtOH.		Solution Buffer	2017-11-20 17:07:37

2. Via dropbox mechanism



Overview of today's openBIS tutorial

Example: RNA sequencing study of 8 different dog breeds

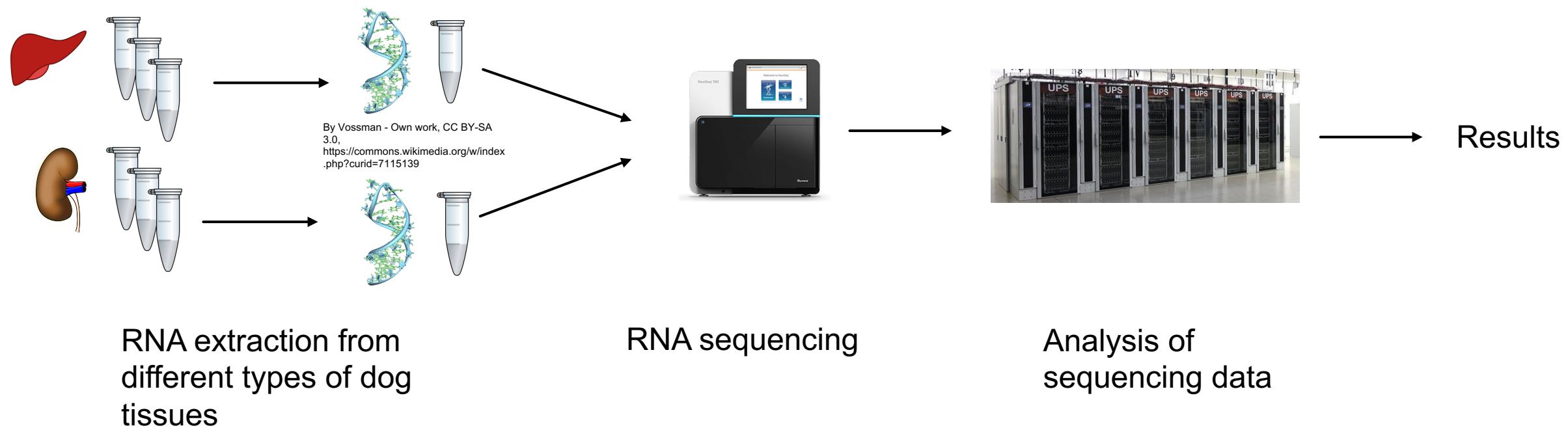
Goal of the study: *understand intra breed genetic variability of dogs with special reference to Beagle dog, since Beagles are used as an animal model for compound testing in the pharma industry.*

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



Overview of the study process

Experiments:

1. RNA extraction
2. RNA sequencing
3. Data analysis

Protocols:

1. RNA extraction
2. RNA sequencing

Samples:

1. Tissues
2. Chemicals
3. RNA extracts



How does the process look like in openBIS?

– Inventory

–  Barillac Materials

–  Samples

 Chemicals

 RNA extracts

 Tissues

–  Barillac Methods

–  Protocols

 RNA extraction protocols

– Lab Notebook

–  My Space (Barillac Barillac)

–  Rna Sequencing Of 8 Dog Breeds

–  RNA sequencing

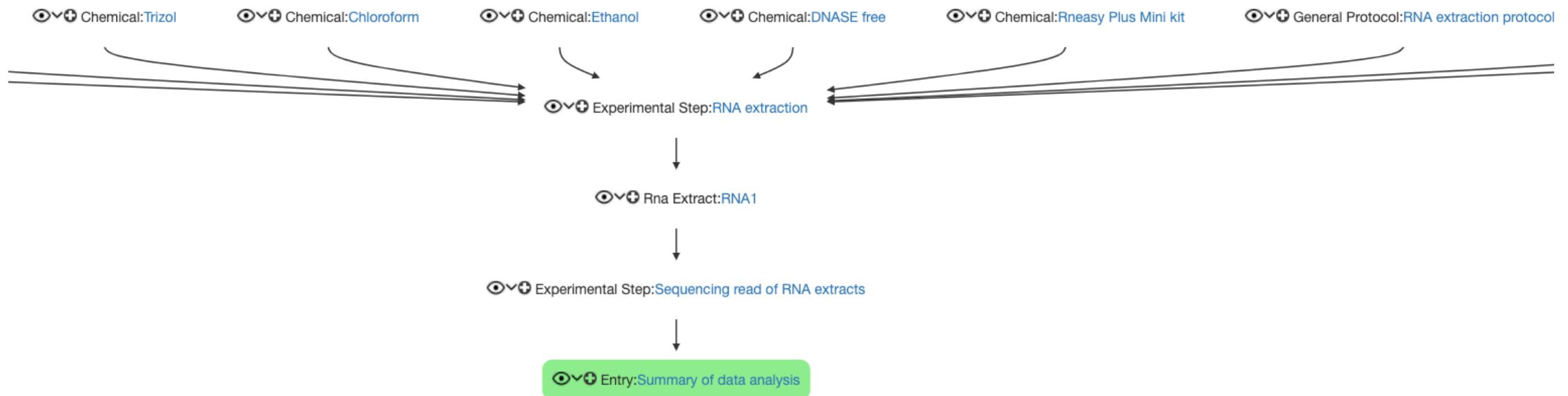
–  RNA extraction

–  Sequencing read of RNA extracts

 RNA seq data

–  Summary of data analysis

How does the process look like in openBIS?

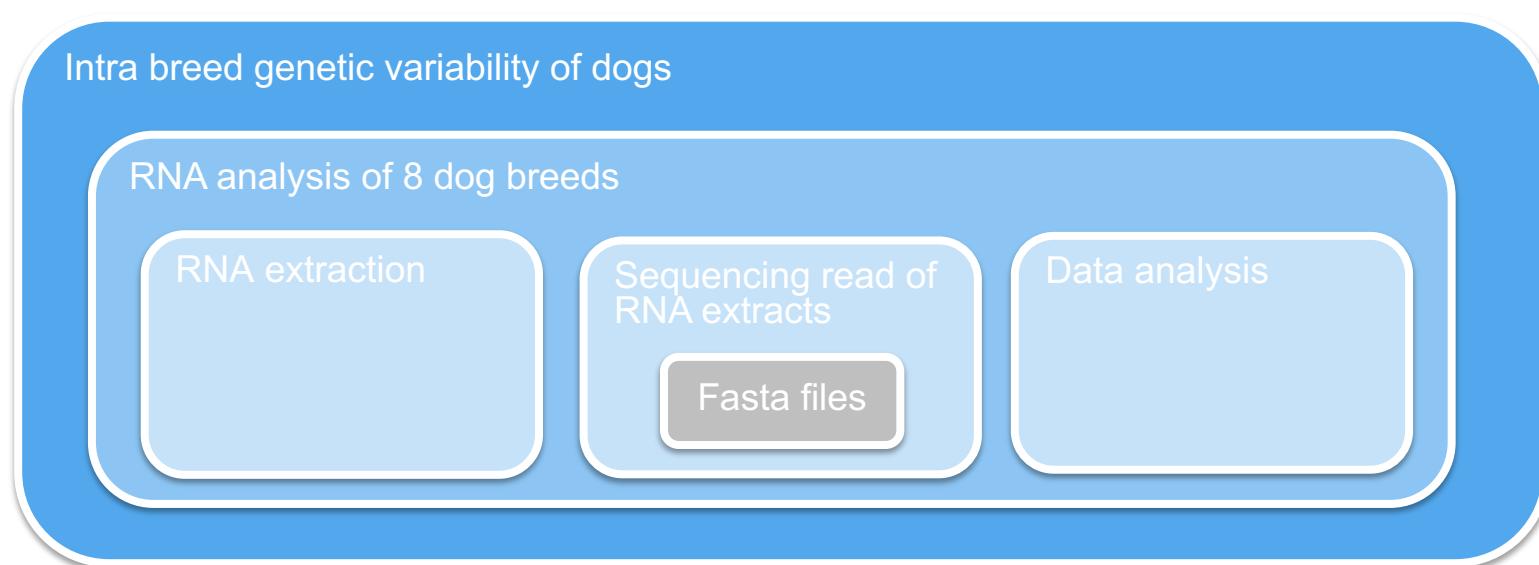


Management of samples and protocols

We will register a few samples and one protocol in the inventory, covering different topics:

1. Single sample registration
2. Batch registration of samples
3. Batch modification of samples
4. How to assign storage positions to samples
5. How to register a standard lab protocol

Lab notebook



- We will create **1 Project, 1 Experiment and 3 Experimental Steps** in your personal space.
- We will see how we can **link** samples and protocols stored in the Inventory to the Experimental Step.
- We will **upload data** to the Experimental Step.

Session 1

- Registration and copy of tissue samples: *sections 3.1.1 + 3.1.2*



10 min

- Inventory
- Barillac Materials
- Samples
 - Chemicals
 - RNA extracts
 - Tissues

Session 2

- Batch registration of chemical samples: *section 3.1.3*
- Deletion of duplicate objects: *section 3.1.4*
- Batch modification of chemical samples: *section 3.1.5*



20 min

- Inventory
- Barillac Materials
- Samples
 - Chemicals
 - RNA extracts
 - Tissues

Session 3

- Visualization of storage positions in the Storage Manager: *section 3.2*

● To do together

Session 4

- Registration of RNA extraction protocol: *section 3.3*



5 min

- ⊖ Barillac Methods
- ⊖ Protocols
 - RNA extraction protocols

Short break



10 min

Session 5

- ➊ Lab Notebook
- ➋ My Space (Barillac Barillac)
- ➌ Intra Breed Genetic Variability Of Dogs
- ➍ RNA analysis of 8 dog breeds
- ➎ RNA extraction

- Registration of project and experiment: section 4.1, 4.2
- Registration of the first Experimental Step: section 4.3.1



15 min

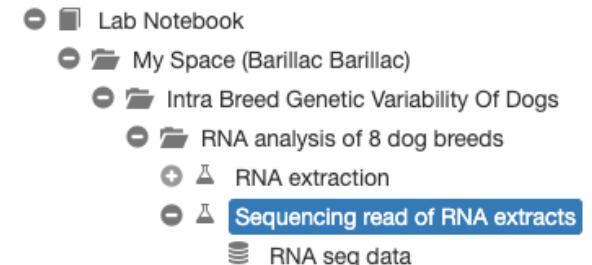
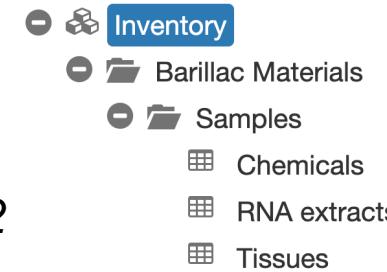
Intra breed genetic variability of dogs

RNA analysis of 8 dog breeds

RNA extraction

Session 6

- Registration of RNA extracts in Inventory: section 4.3.2
- Registration of RNA sequencing Experimental Step + data upload: sections 4.3.3, 4.3.4



10 min

Intra breed genetic variability of dogs

RNA analysis of 8 dog breeds

RNA extraction

Sequencing read of
RNA extracts

Fasta files

Session 7

- Data visualization: *sections 4.3.5*
- Data & metadata exports: *4.3.6*
- To do together

Session 8

- Registration of Data Analysis Entry: *section 4.3.7*
- User access rights: *section 4.4*

- Lab Notebook
- My Space (Barillac Barillac)
- Intra Breed Genetic Variability Of Dogs
- RNA analysis of 8 dog breeds
- RNA extraction
- Sequencing read of RNA extracts
- RNA seq data
- Summary of data analysis



10 min

Intra breed genetic variability of dogs

RNA analysis of 8 dog breeds

RNA extraction

Sequencing read of
RNA extracts

Fasta files

Data analysis

Session 9

- Searching the ELN: section 5 (5.1, 5.2)



10 min

Session 10

- Freezing entities: *section 6*
- Jupyter notebooks: *section 7*
- To do together

Paperplane



New service for keeping track of published data offered by SIS:

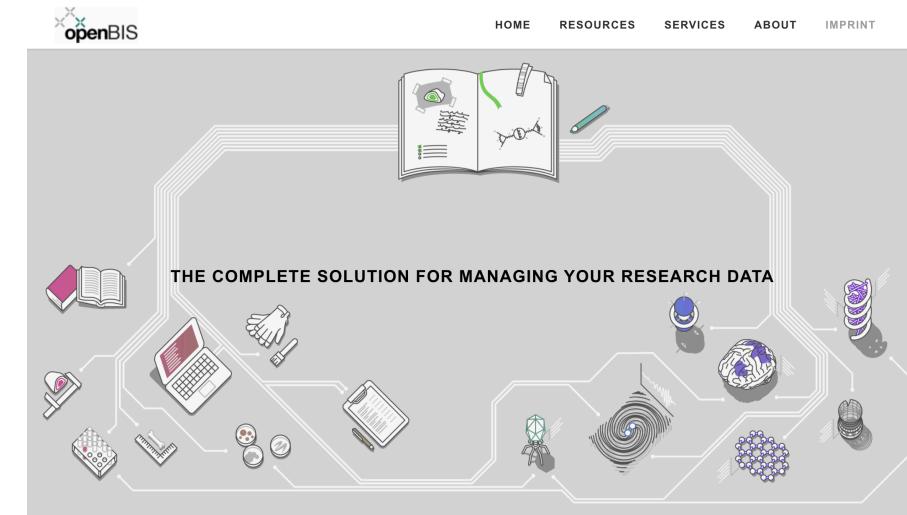
<https://www.paperplane.ethz.ch>

Contacts & useful info

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

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

Twitter: https://twitter.com/ETH_SIS



SIS helpdesk

sis.helpdesk@ethz.ch

Caterina Barillari

caterina.barillari@id.ethz.ch

Priyasma Bhoumik

priyasma.bhoumik@id.ethz.ch