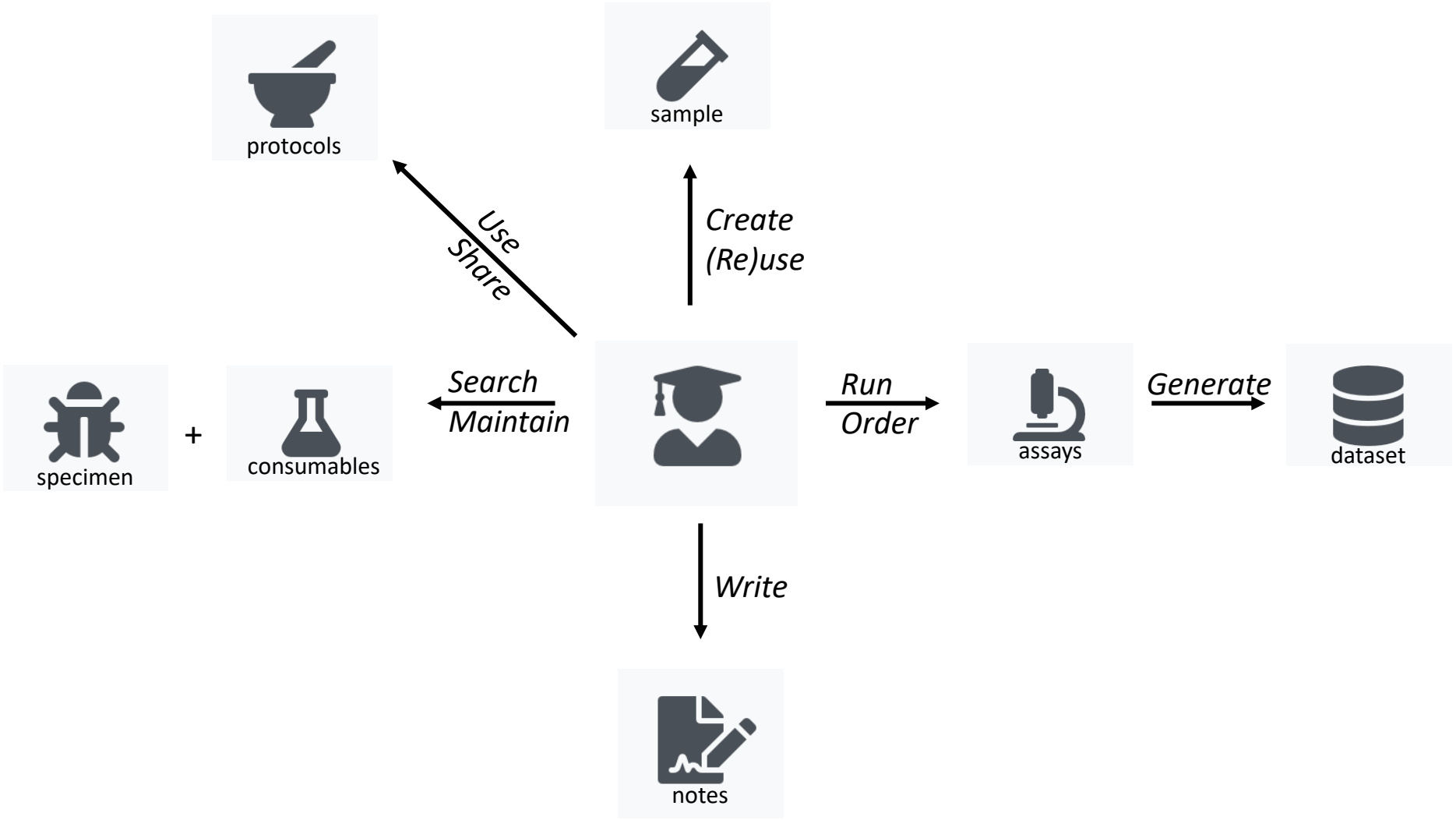
STOCKS Training: Sample Management

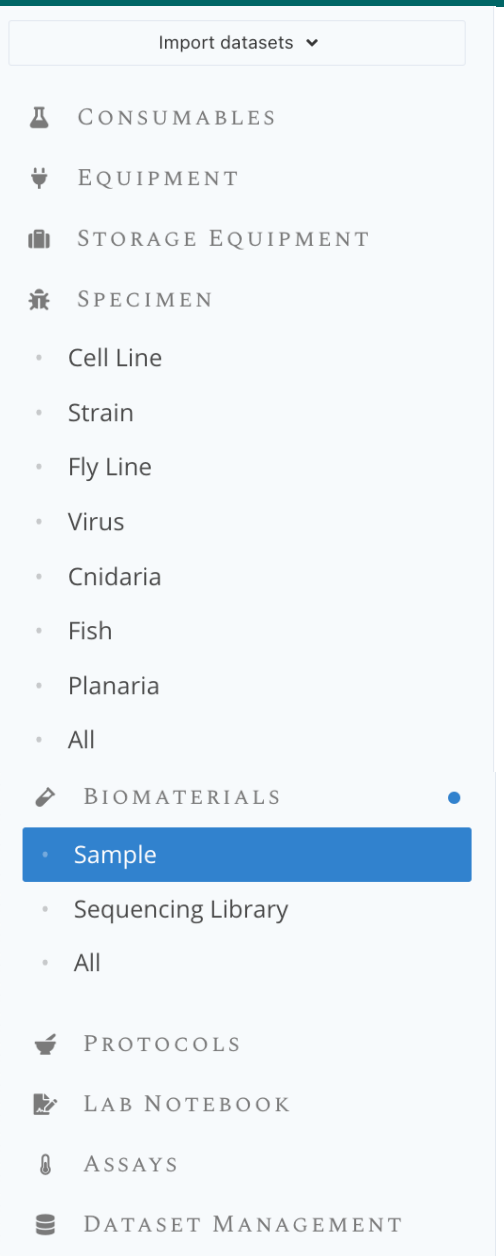
Charles Girardot

Genome Biology Computational Support (GBCS)

Samples are consumed and created everyday to generate data



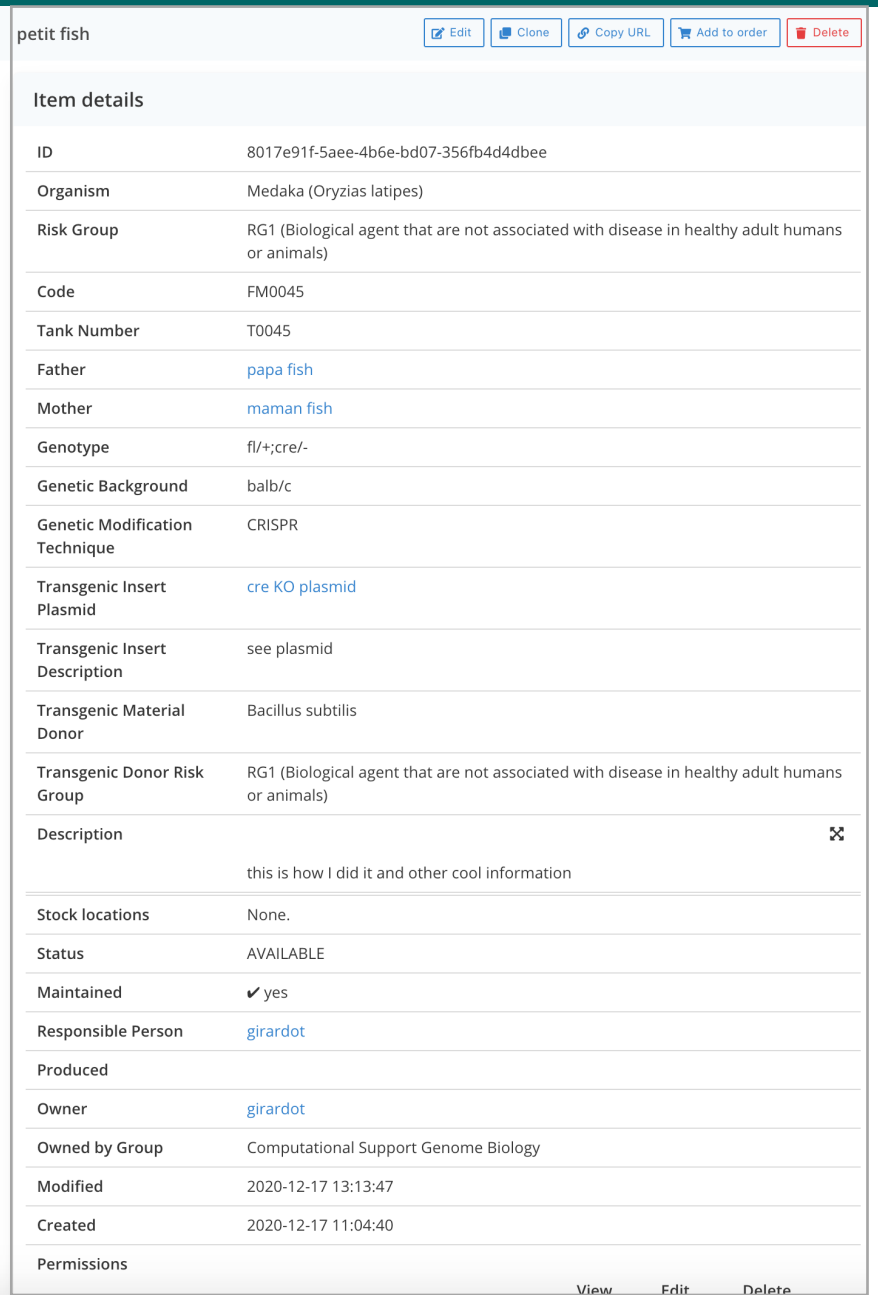
The sample management module



- **Specimen**: cell line, fly lines, strains…
  - The living biomaterial from which your samples are coming from
    - **Parent-child relationship**
  - More **specimen types** can be easily **added**
  - **Additional properties** can easily be **added**

- **Biomaterial (aka sample):**
  - **Consumed** and **produced** during your **experiments**
  - The **Sample model** is the **base** model fitted for most needs
  - **Sequencing Library** extends the base sample and add options specific to **sequencing assays**
  - New **sample sub-types** require **addition of code**
  - Missing properties can be added to existing types

Specimen are very similar to the lab inventory collections



- Similar features as for the consumables / equipment collections

Properties to parents can be defined when relevant

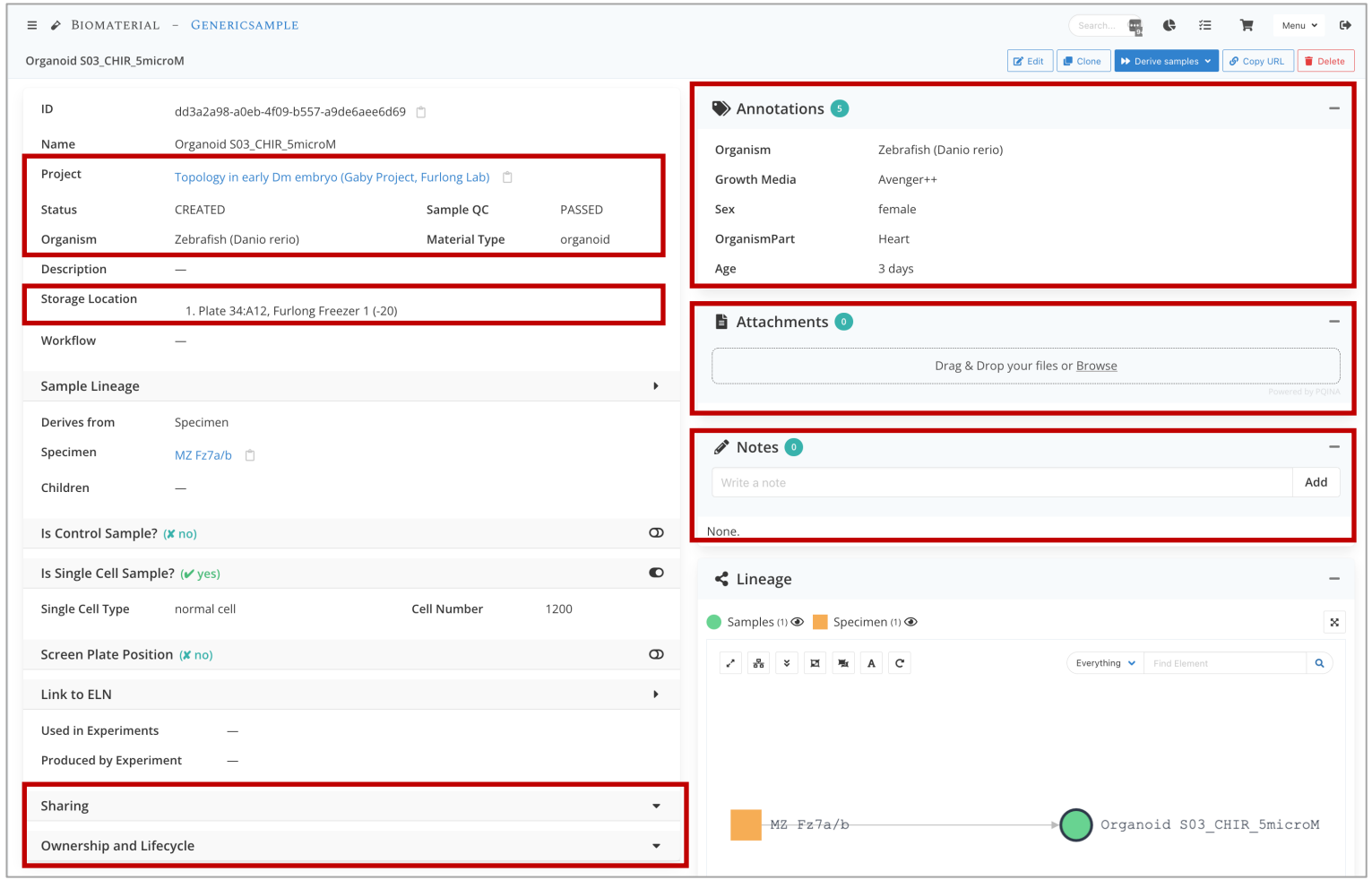
GMO-related properties to support (German) regulation needs

The **Sample** biomaterial has the usual properties

Primary project & organism are mandatory

Storage Location(s) available

Sharing & ownership



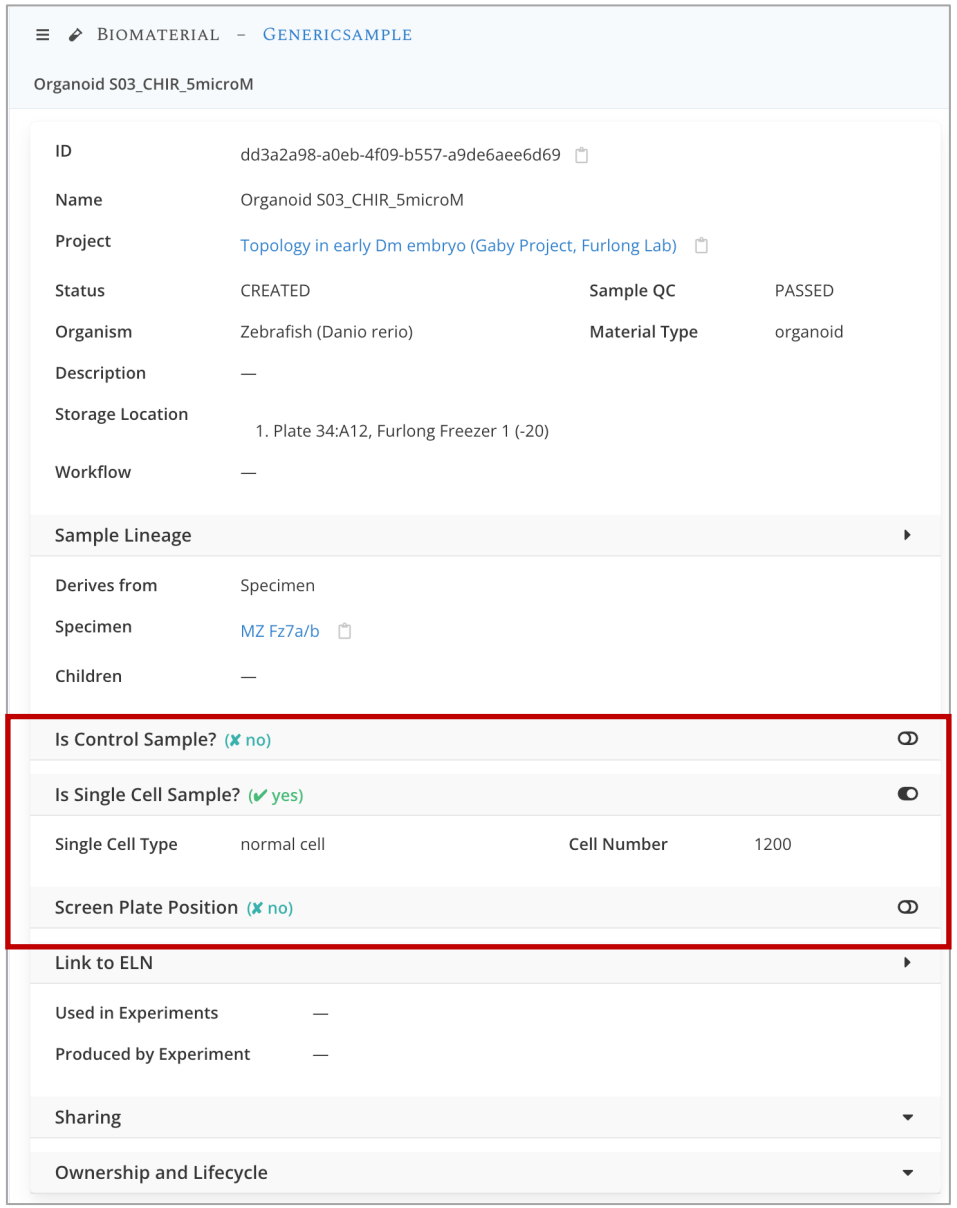
Annotations

Attachments

Notes

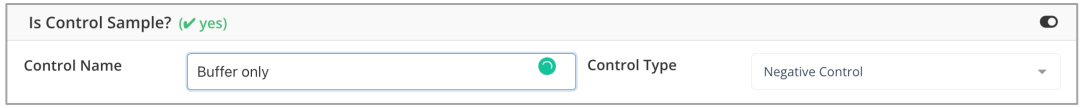
The **Sample** biomaterial should fit most use-cases

Activable sections



- Designed for supporting array-based screens

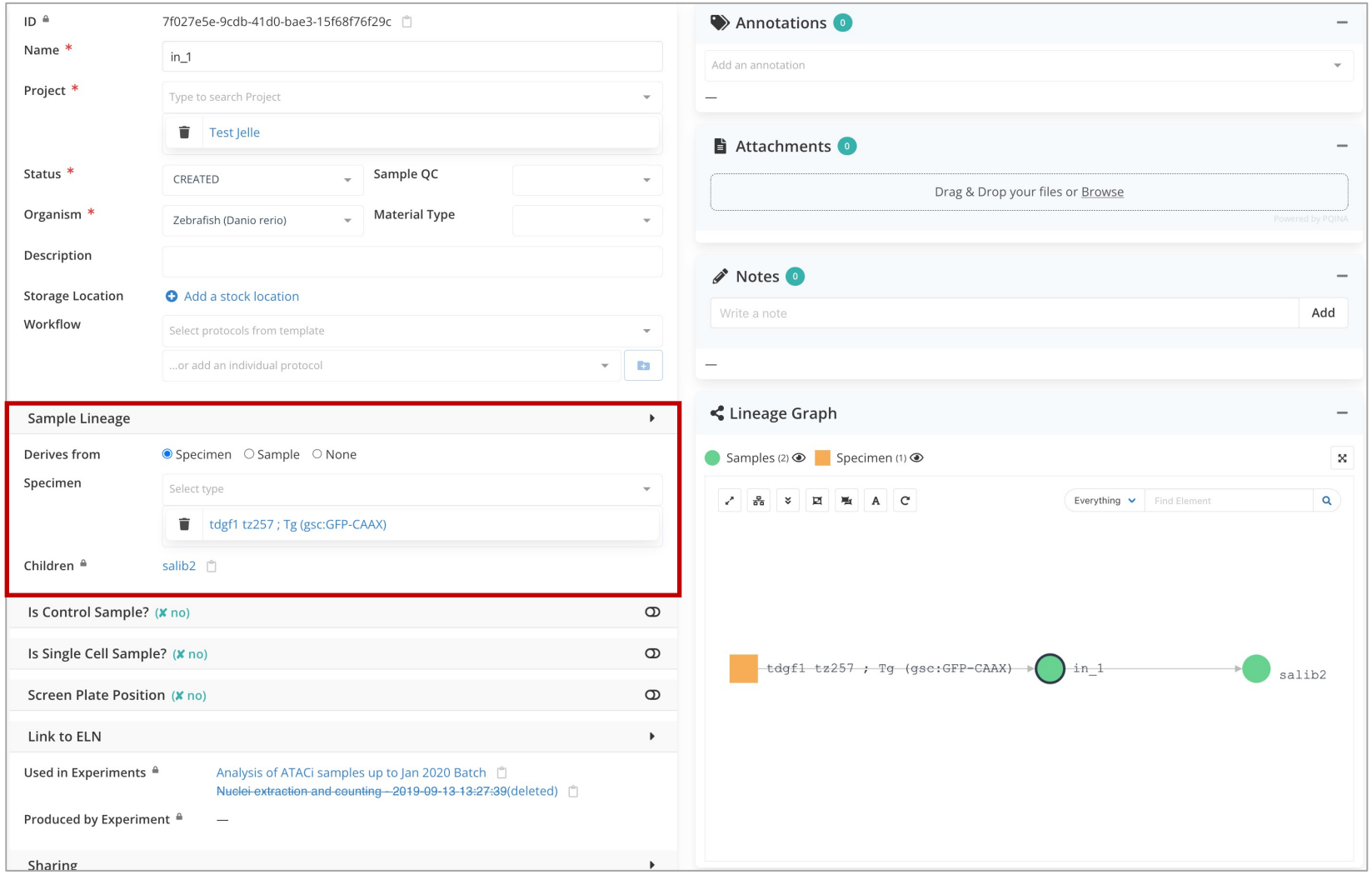
Support for control samples



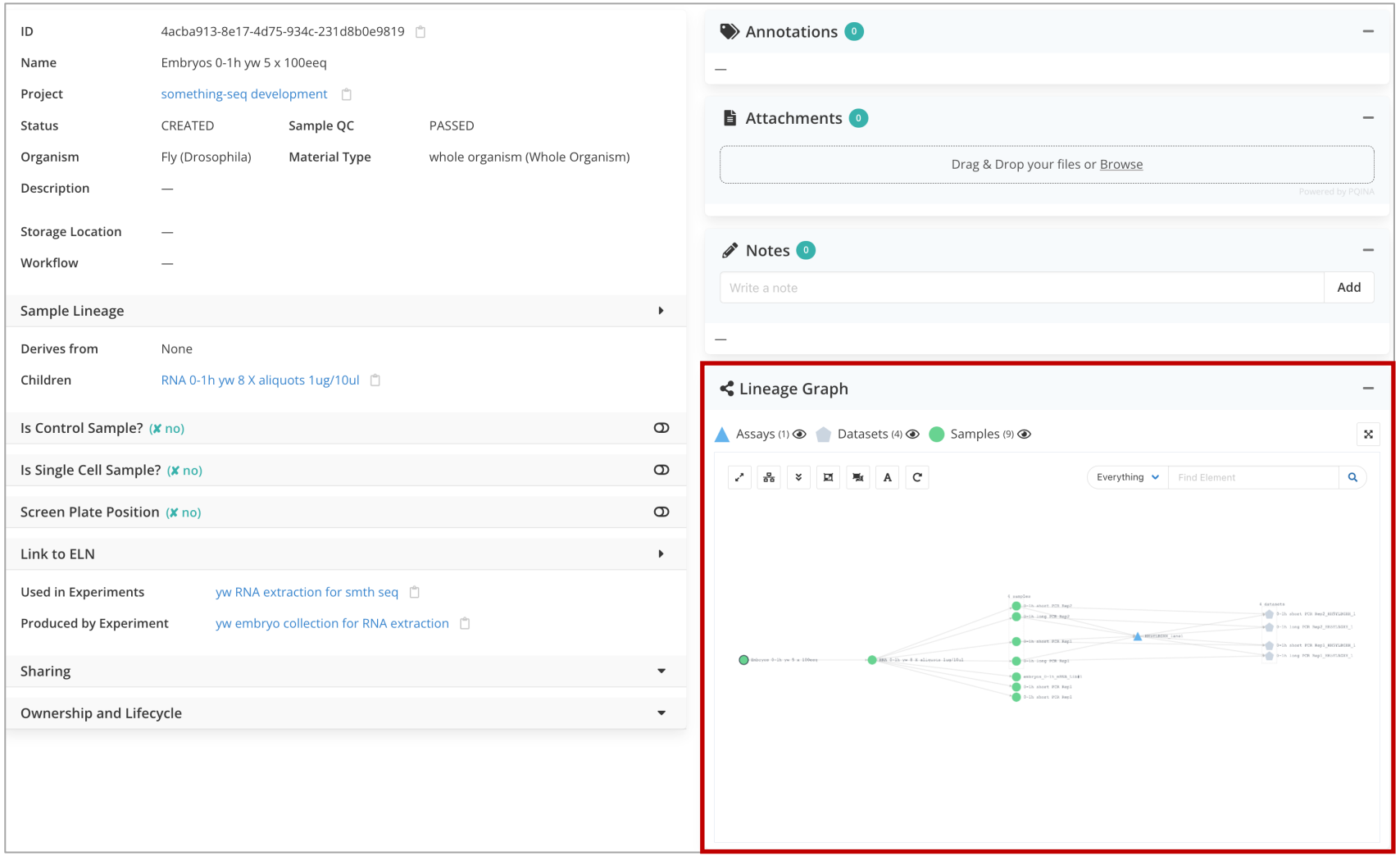
Support for plated samples

Samples can have parent sample(s) or specimen

Either:
- One Specimen; or
- One or more Samples



The lineage reflects the complete experimental procedure

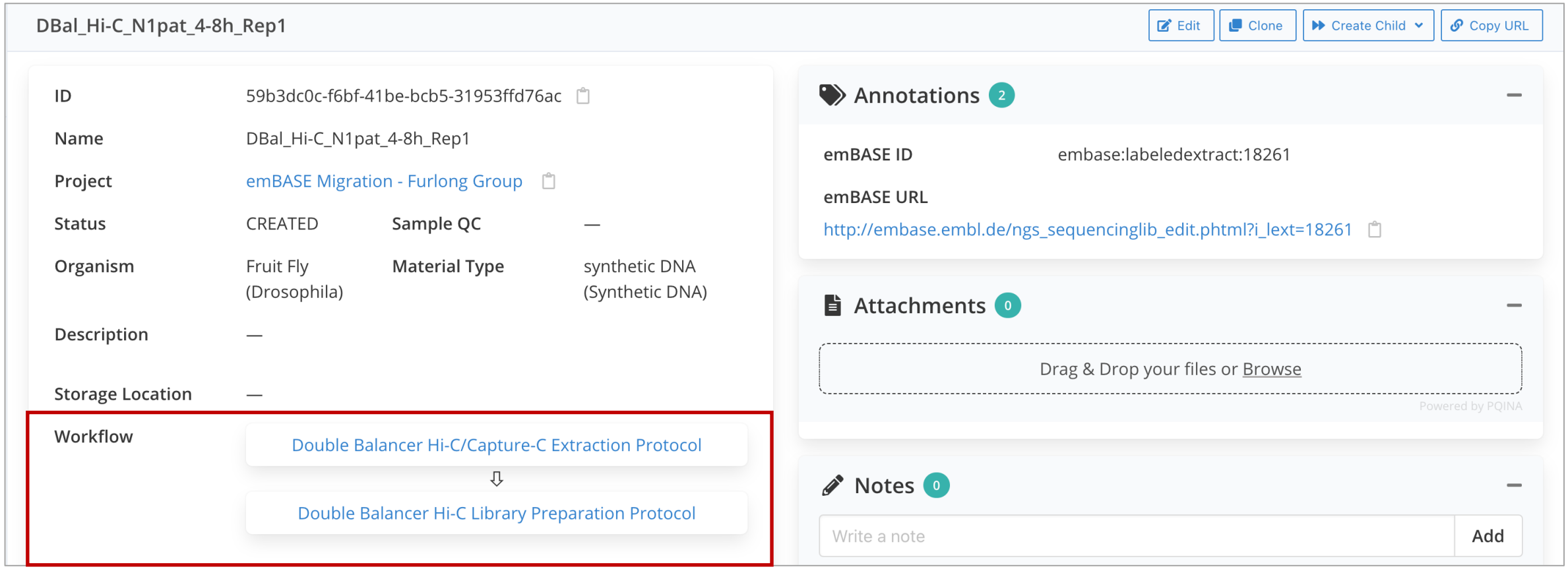


Embedded viewer with many display options and name search

✓ As many steps as you need
✓ Sample Merge allowed

Samples must indicate how they were generated

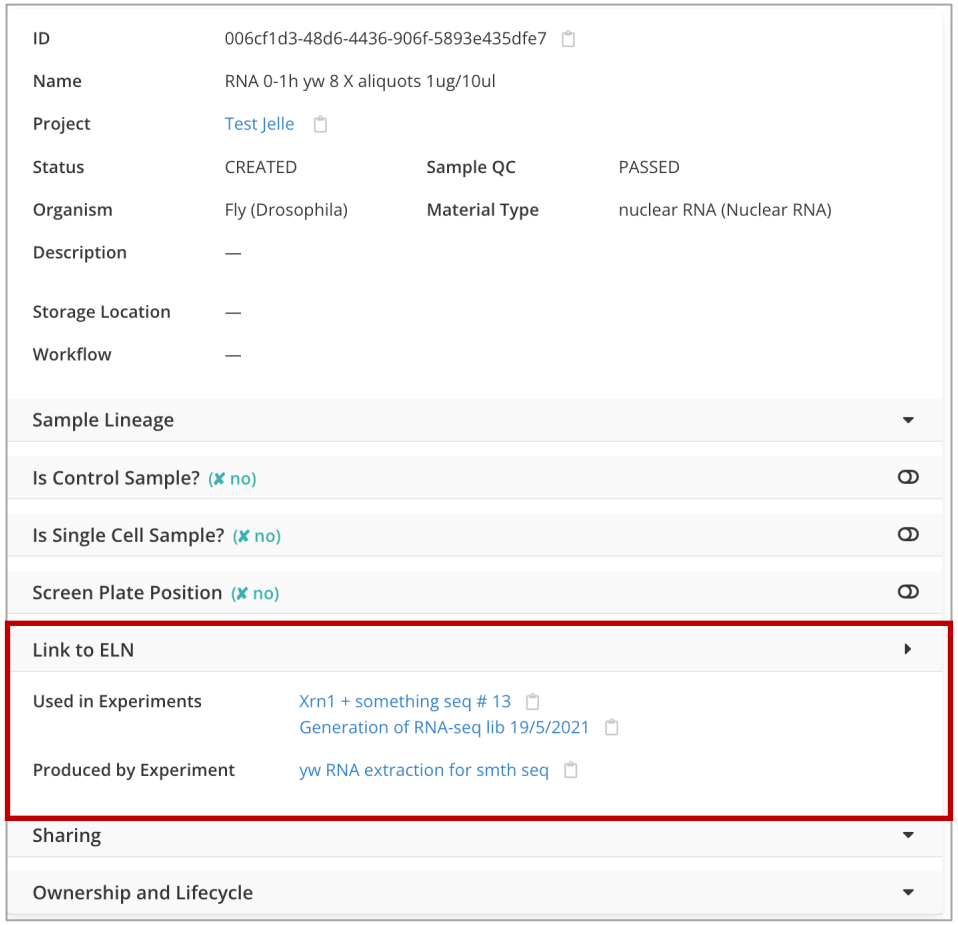
- Workflow = [Protocol1, Protocol2, … , ProtocolN]
- Reflects how the sample was generated from the immediate parent



➢ Needed at publication time

Samples are consumed and generated by experiments

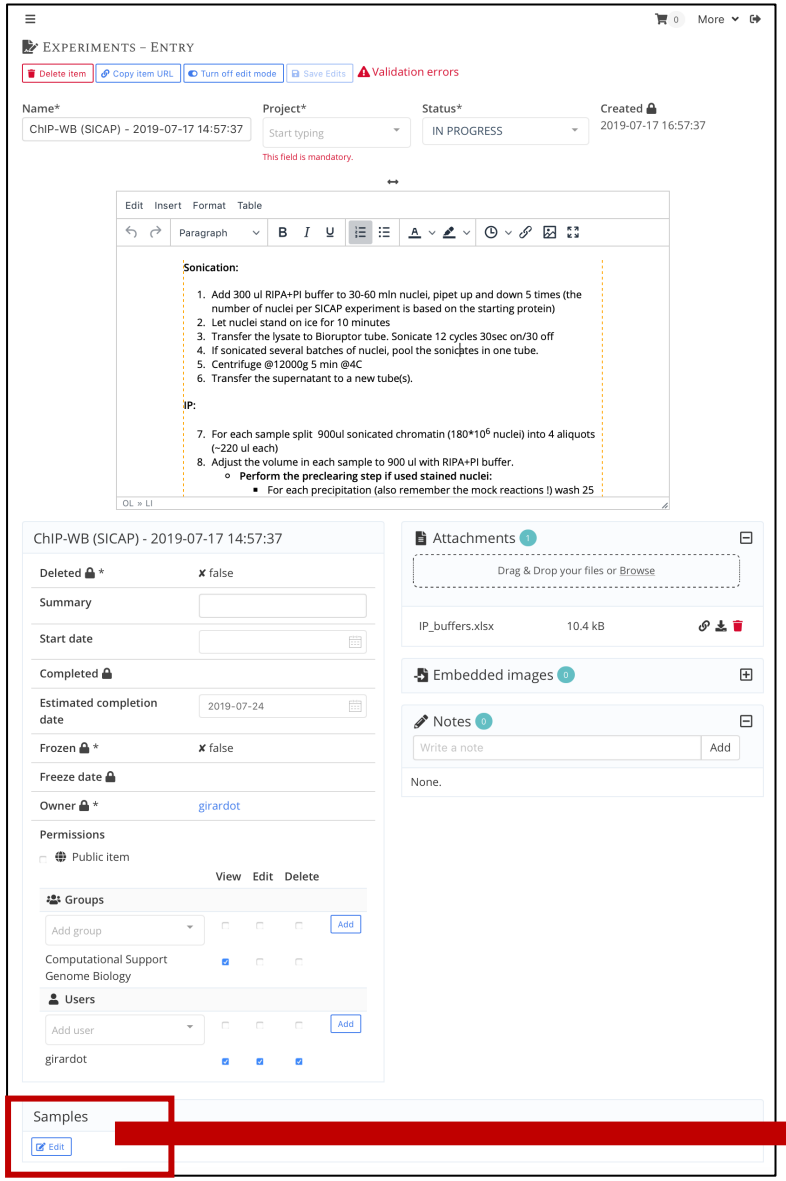
➢ Samples are connected to your ELN's *experiments*

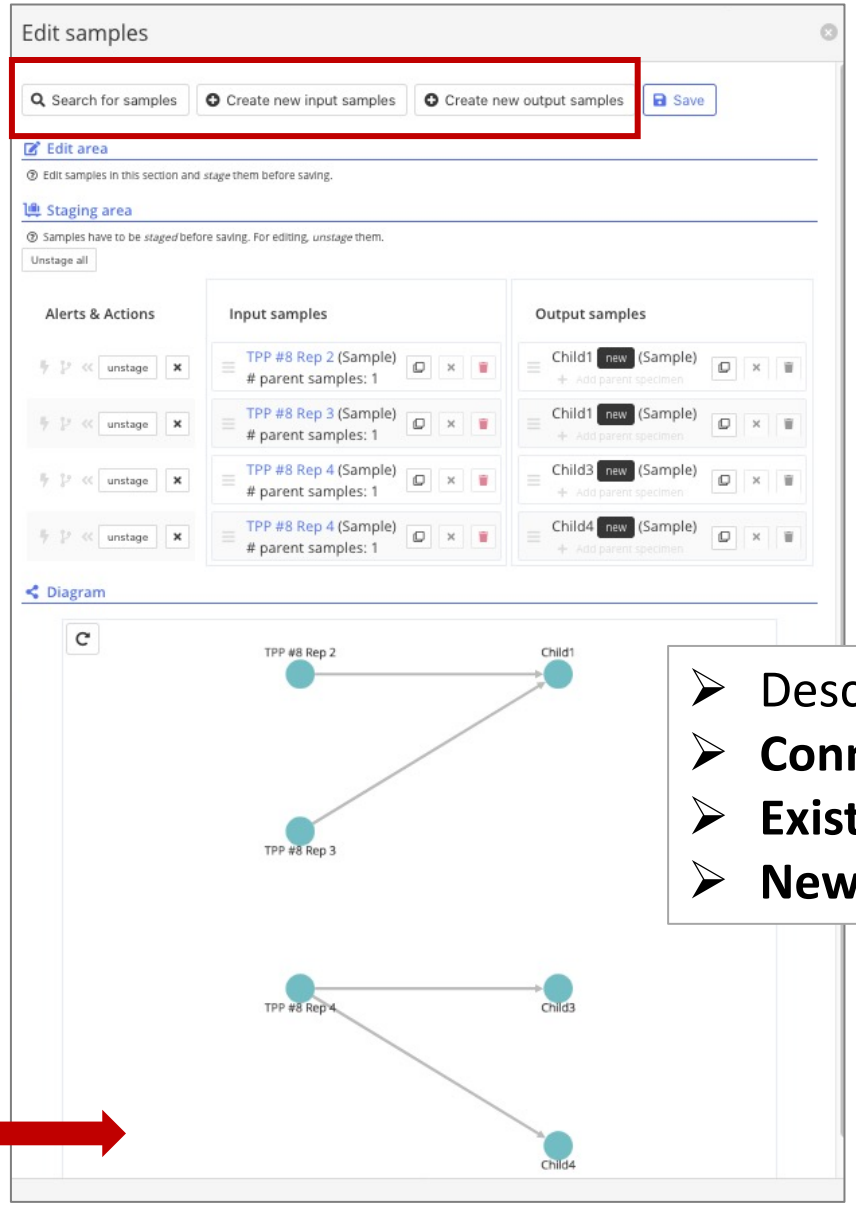


- Samples may be consumed or "used" in multiple experiments
- Samples are produced by a unique experiment

➢ Connecting samples to experiments happens <u>on the Experiment page</u>

Link up Samples & Experiments with the Sample Editor

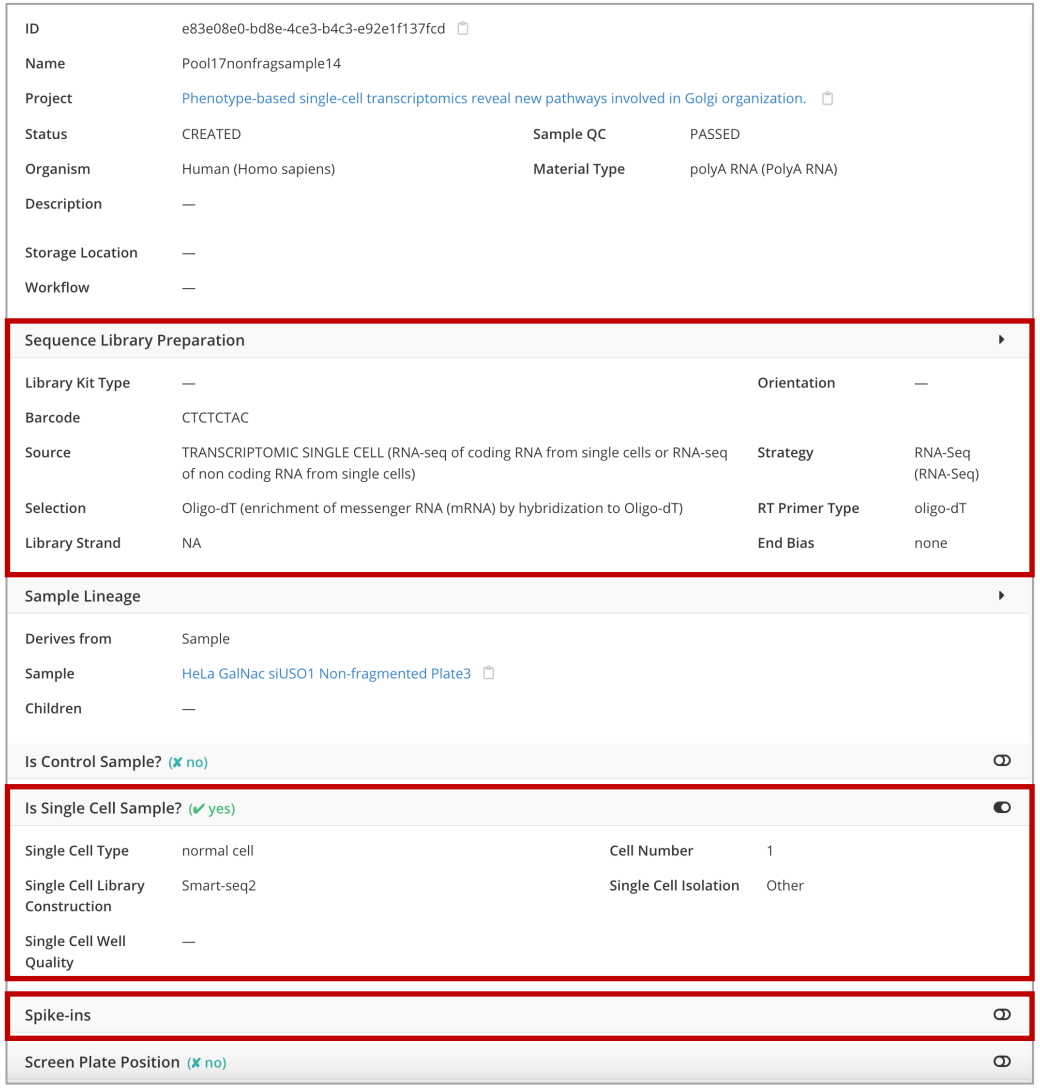




- Describe the ***input*** and ***output*** **samples**
- **Connect** samples (parent-child)
- **Existing samples** can be used
- **New samples** can be batch created

The Sequencing Library fits the Sequencing Assay needs

- The Sequencing Library = the base Sample + sequencing library-specific attributes



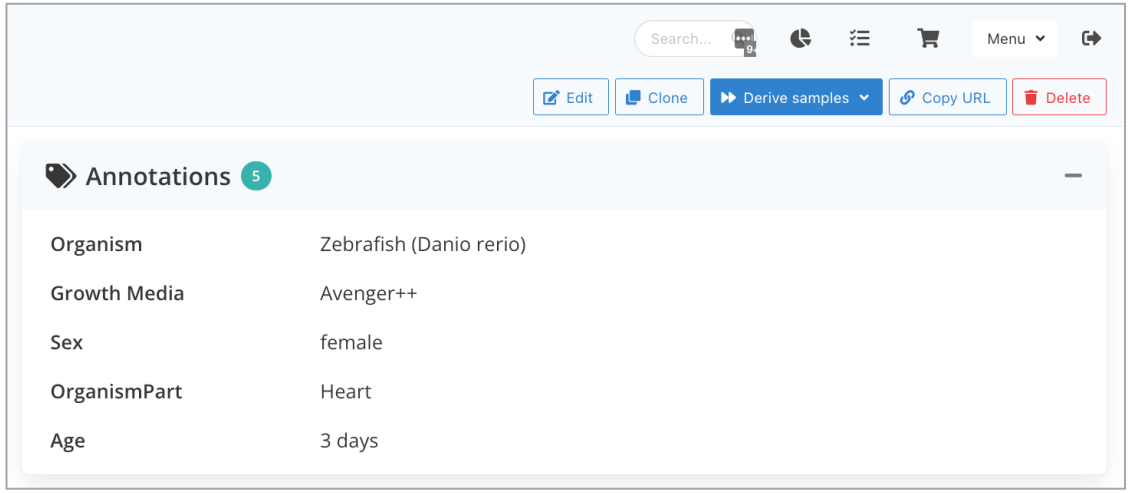
Sequencing library preparation section

- Key aspects of the library preparation procedure
- **Required at submission time**
- Values match those required by public repositories

Single-cell sample section augmented with properties required by public repositories

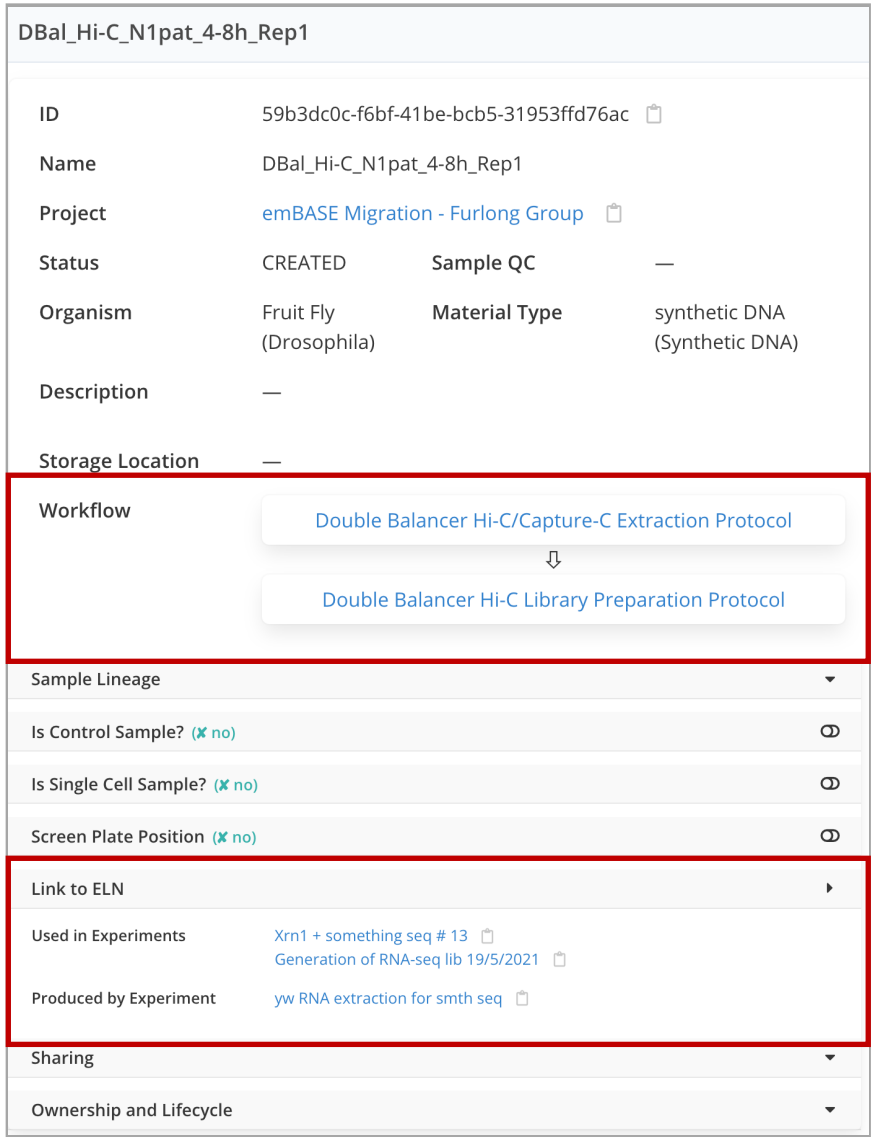
Additional spike-ins section to reflect common NGS design

Samples must be annotated to be FAIR



- Annotations reflect both the project and the sample specificity
- Reflect general sample characteristics e.g. cell line, cell type, organism, sex, genotype, …
- Define important experimental factors i.e. reflecting the study design
  - How do samples group together e.g. treatment, phenotype, targeted epitope (ChIP) ?
- Biological annotations are mandatory to deposit data in public repositories

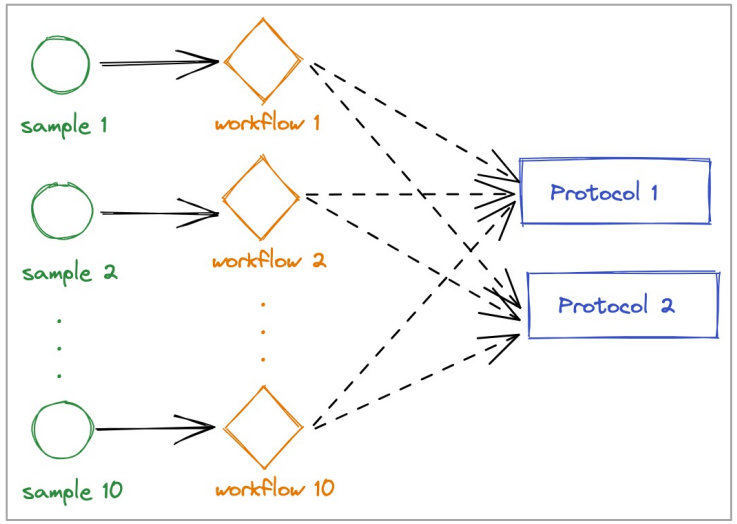
Linking samples to experiments and protocols is not redundant



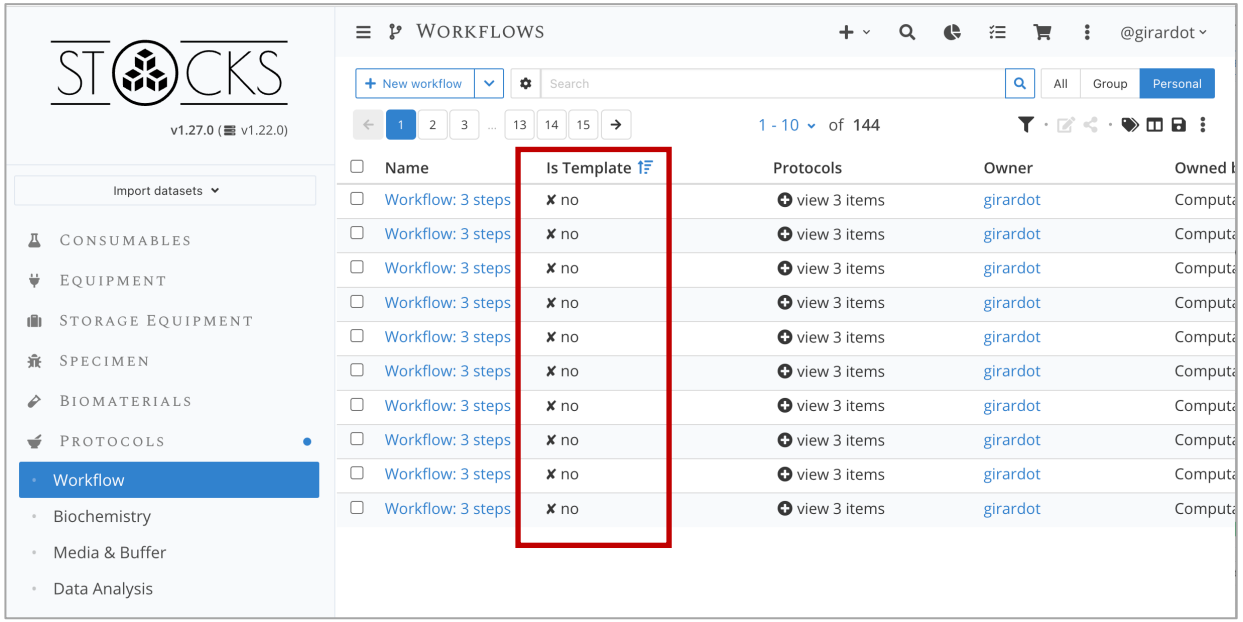
- Samples can be linked to both protocols and ELN's experiment
  - *Not everyone uses ELN*
- Samples **must** link to protocols
  - At data deposition time, we submit protocols not your ELN's notes
  - In fact we even need a protocol summary (no structured text)
    - Protocols have a *summary* property for this

Each sample has its own workflow

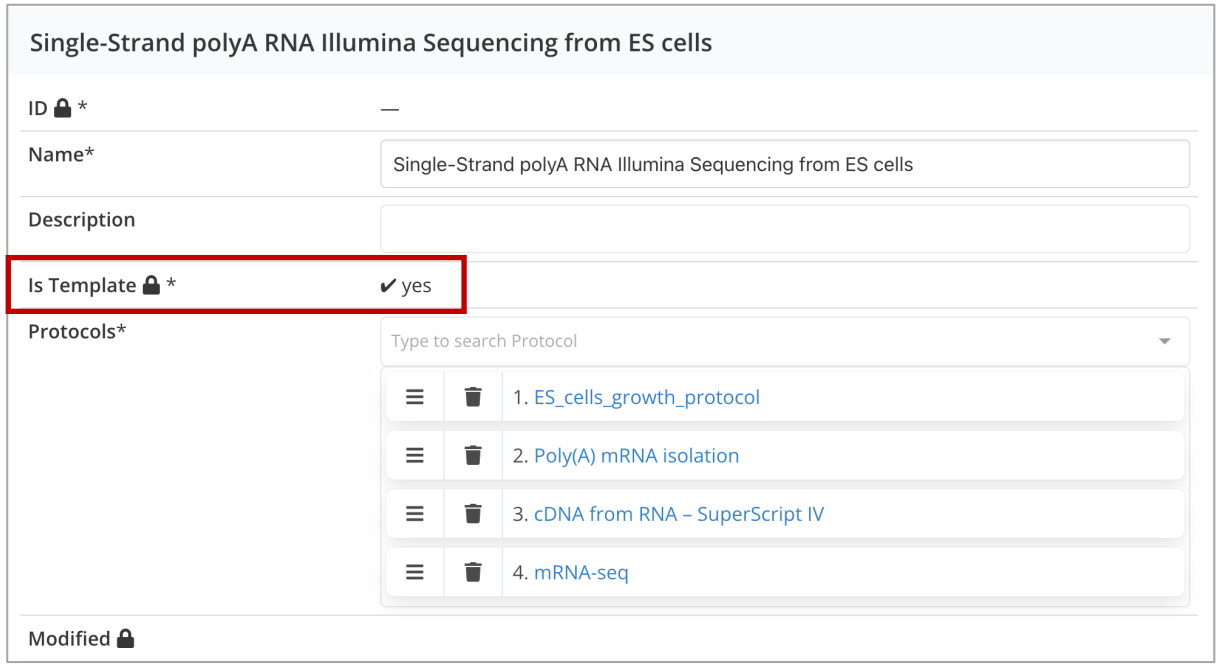
- When adding protocols to a sample, you modify its workflow
- When multiple samples link the same protocols, their workflows are still distinct



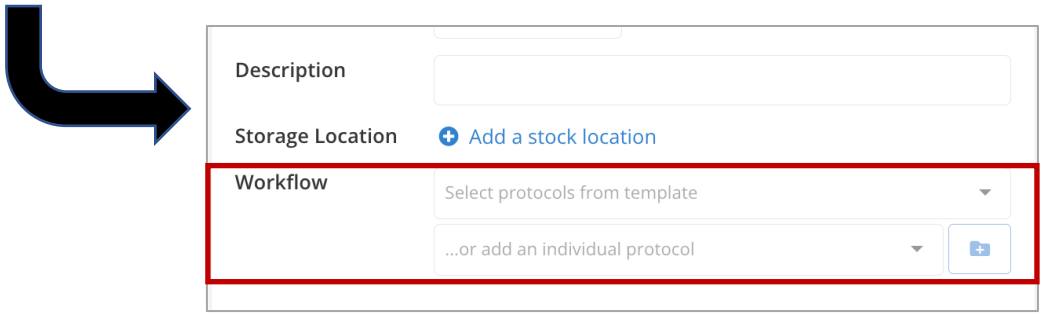
- WFs are visible in the Workflow menu
- WFs are named identically
- WFs can only be edited from the sample page



Workflow templates can be created for common procedures

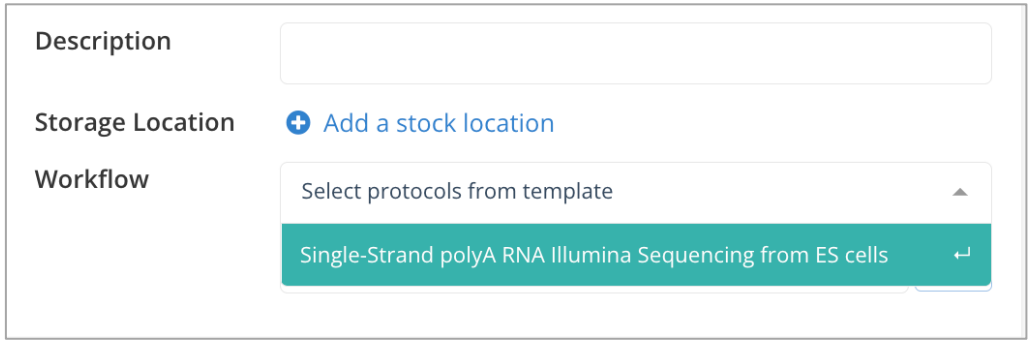


- Create a workflow template reflecting a chain of protocols often used (e.g. core facilities)



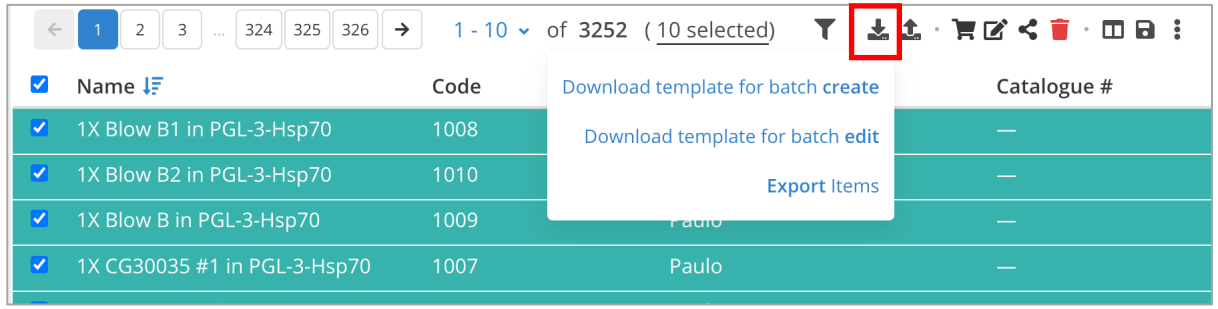
Sample Page offers 2 options under workflow

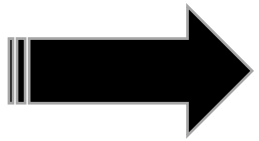
- We used the second one before



Sample Page

Samples can also be batch-created and edited





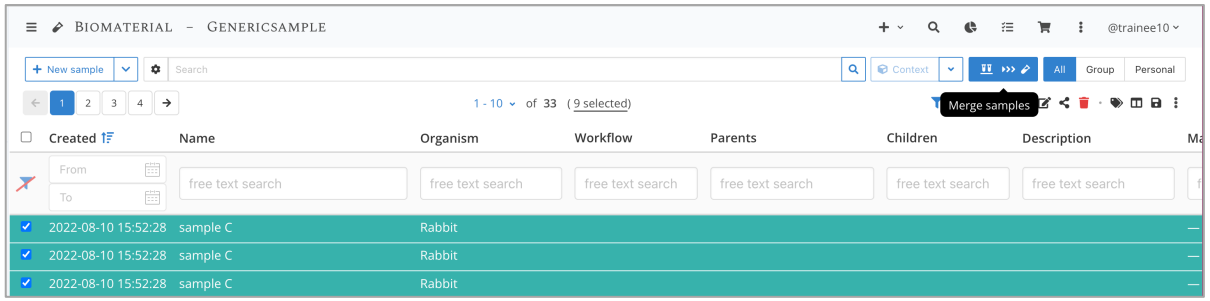
- Works as for any lab inventory collection
- Few additional tricks to know
  - The **ID** of the **Parent(s) and Protocol(s)** must be provided, not their name
  - When providing multiple protocols, their order is preserved in the workflow

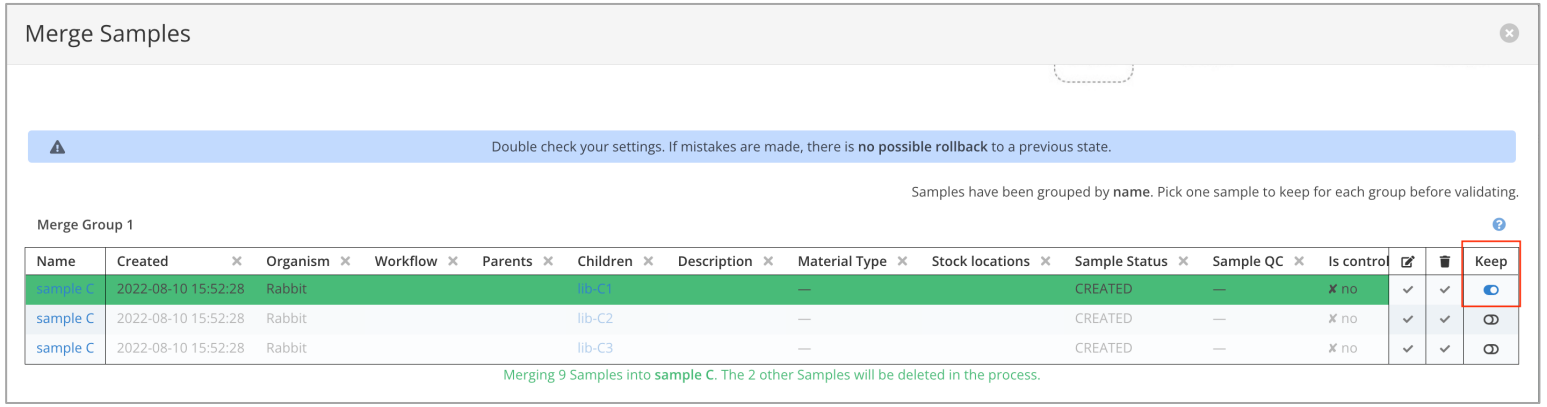
Sample can be registered in different ways …and duplicated

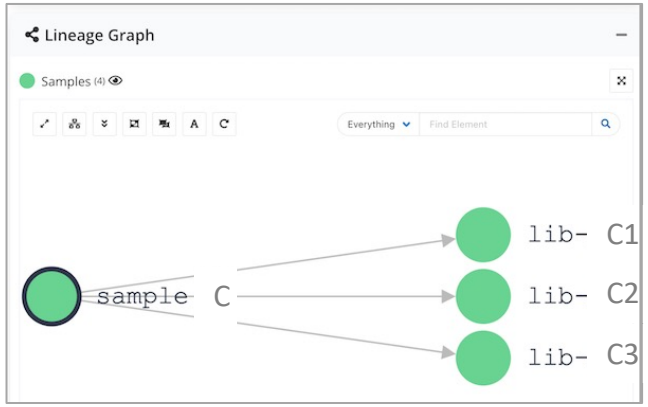
1. Manually from the list page
2. With the Experiment's sample editor
3. In batch using excel-templates
4. *At Dataset Registration*
   - similar to the experiment's sample editor

➢ Samples get duplicated for many reasons
- You create a new one instead of using the existing one
- Your colleague already created the same sample
- The sample was existing but not shared with you
- …

Getting rid of redundant samples



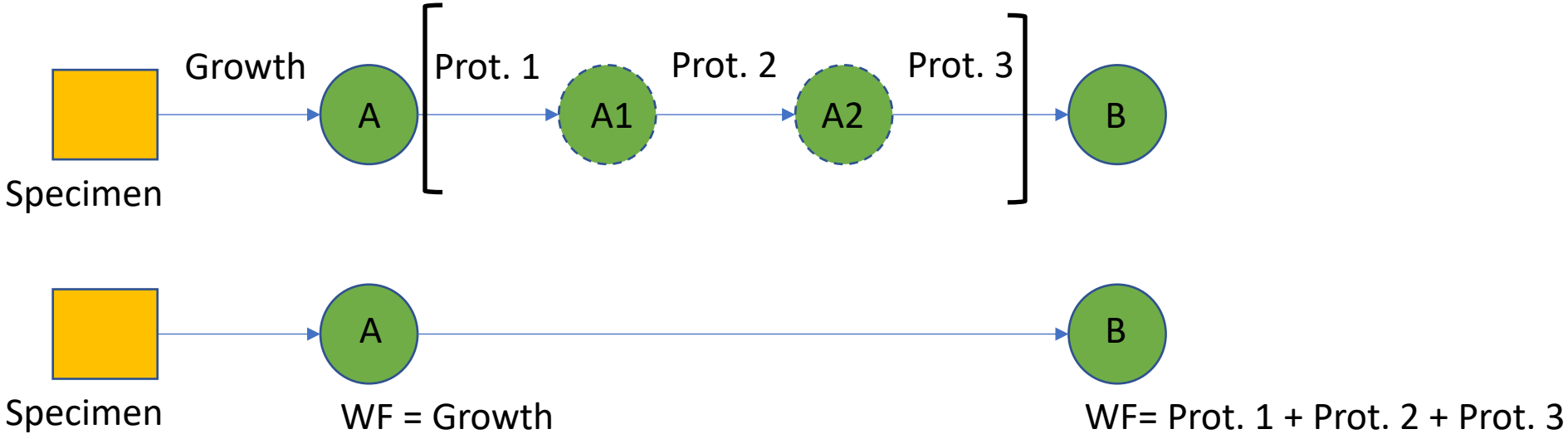




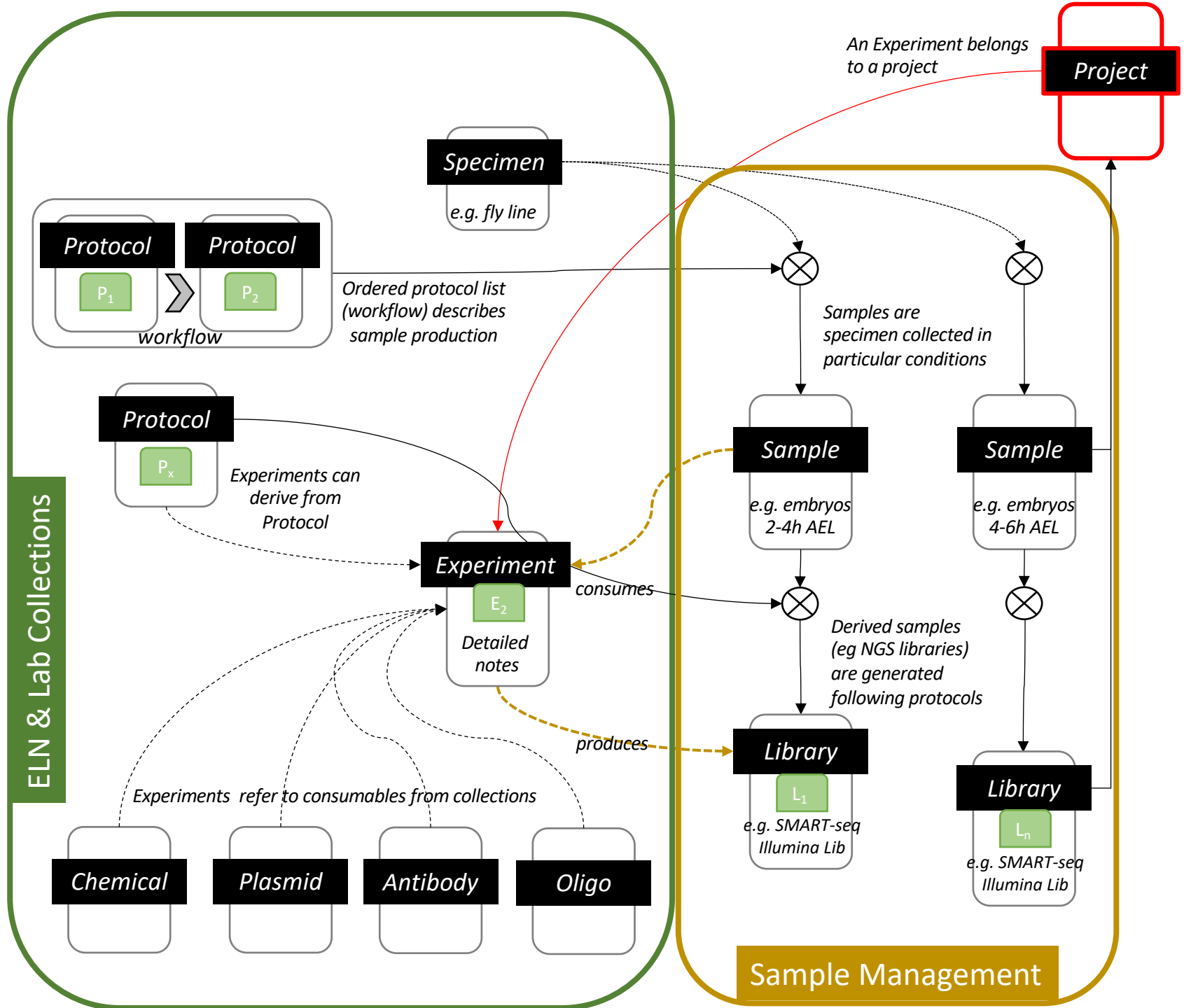
- One sample is kept, the others deleted
- Parent-child relationships are migrated to the kept sample

When should samples be created?

- register your samples as they are generated
  - when taking your notes
- usually not needed to reflect transient samples
  - create samples when found in storage and can be re-used



- To reflect experimental design aspects (parallel processing using split-merge…)


An Experiment belongs to a project
Project
Specimen
e.g. fly line
Protocol
Protocol
P1
P2
workflow
Ordered protocol list (workflow) describes sample production
Samples are specimen collected in particular conditions
Protocol
Px
Experiments can derive from Protocol
Sample
e.g. embryos 2-4h AEL
Sample
e.g. embryos 4-6h AEL
Experiment
E2
Detailed notes
consumes
Derived samples (eg NGS libraries) are generated following protocols
produces
Library
L1
e.g. SMART-seq Illumina Lib
Library
Ln
e.g. SMART-seq Illumina Lib
Experiments refer to consumables from collections
Chemical
Plasmid
Antibody
Oligo
ELN & Lab Collections
Sample Management

Thank you for your attention

Jelle Scholtalbers

Matthias Monfort

Nayeem Reza

Questions ?