

**Luca Bertinetti**  
Chitin based materials and tools group

# First steps towards effective data management in a multidisciplinary research group

27.05.2024 — From Paper to pixels CMCB/PoL Campus Workshop

# Outline

Data management in an interdisciplinary group... WAM!! (What A Mess)

A possible strategy to deal with it

Our implementation using eLabFTW

# Variety of experimental tools

## Raw data generation

Wet chemistry synthesis/purification  
LM (from usb camera to CLSM and FLIM)  
SEM - TEM - FIB/SEM @ RT and cryo  
X-ray scattering (synchrotron)  
X-ray tomography  
EM tomography  
AFM and force spectroscopy  
Nanoindentation  
Made up setups  
Vibrational spectroscopy (Raman and IR)  
Water sorption calorimetry + TG/DSC  
Proteomics

## Data processing —> further data generation

Bioinformatic  
3D Image processing  
3D Image segmentation  
3D Texture recognition  
Multimodal data reduction  
Hyperspectral data processing  
Advanced fitting  
Quantitative 3D morphological and orientational analysis  
...

# Variety of sample types

Appendages and materials belonging to specific parts of arthropods (hairs, scales, legs and tools, sensory organs ...)

Cells, tissues and organisms (living or fixed)

Solutions/dispersions of polysaccharides and proteins

...

# The unFAIR nightmare

End up with huge amounts of data storage we don't know what they are

# The pressing need

Store some minimal information together with the data

# The FAIR dream

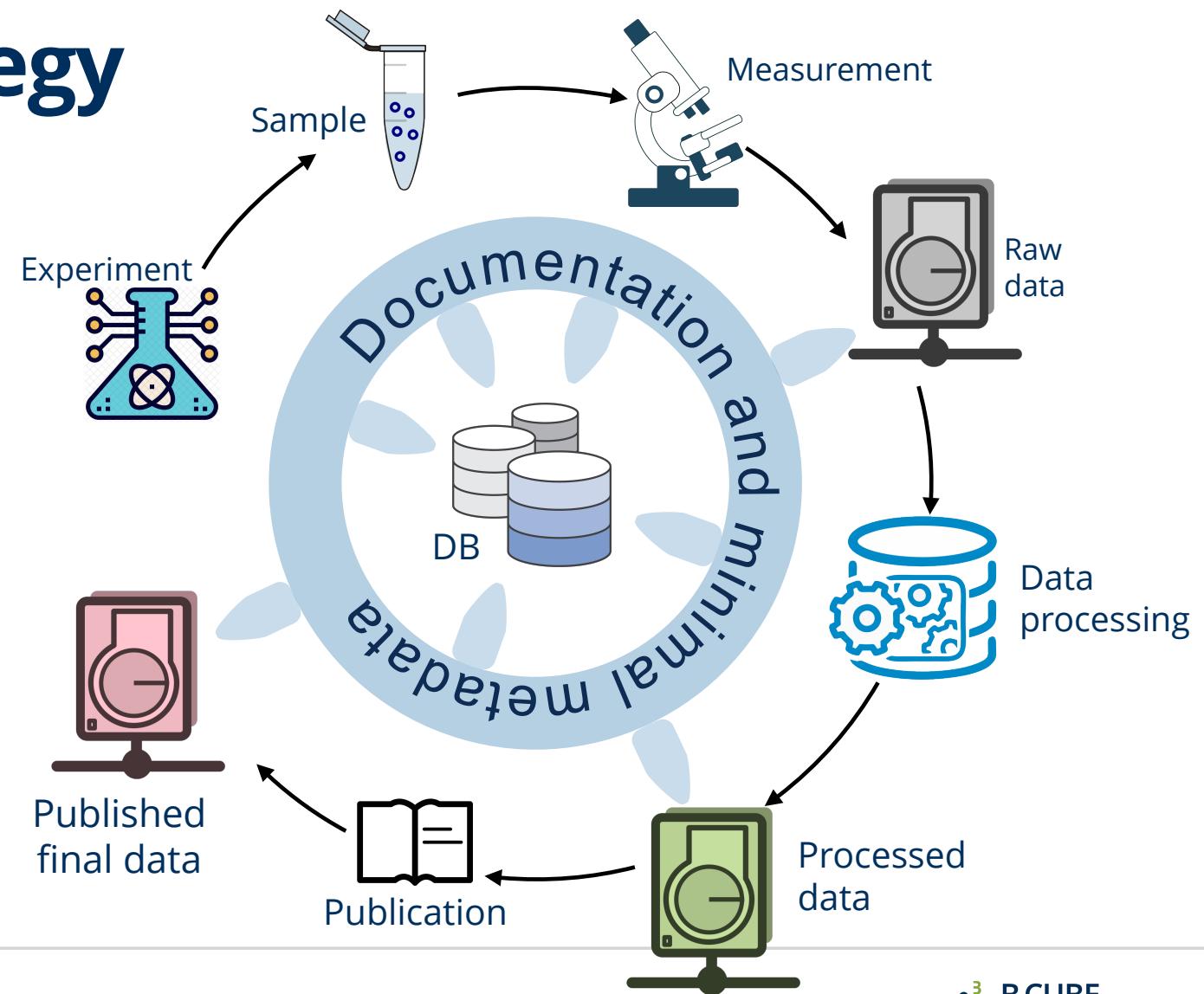
Minimal information to be linked to a thorough documentation

# A possible strategy

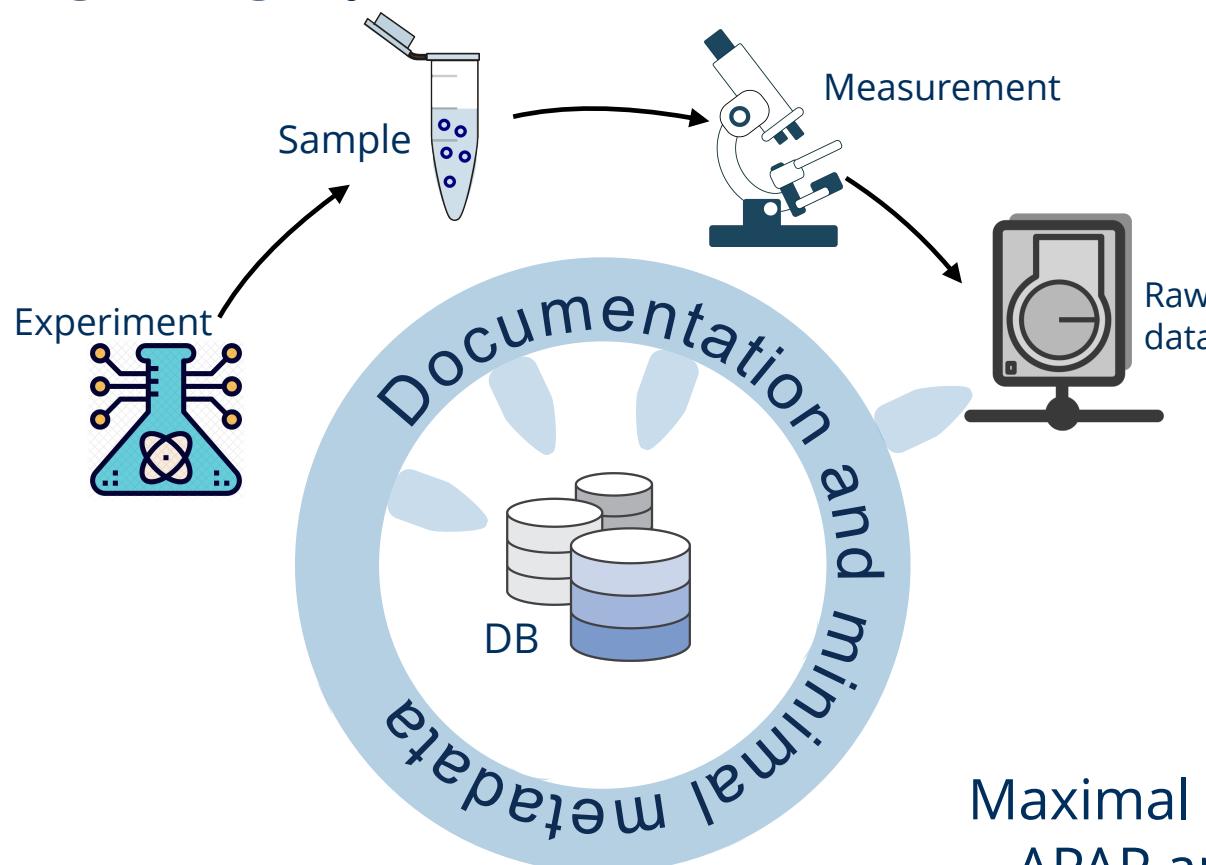
Management tool (MT)

User input to MT

MT output to network share



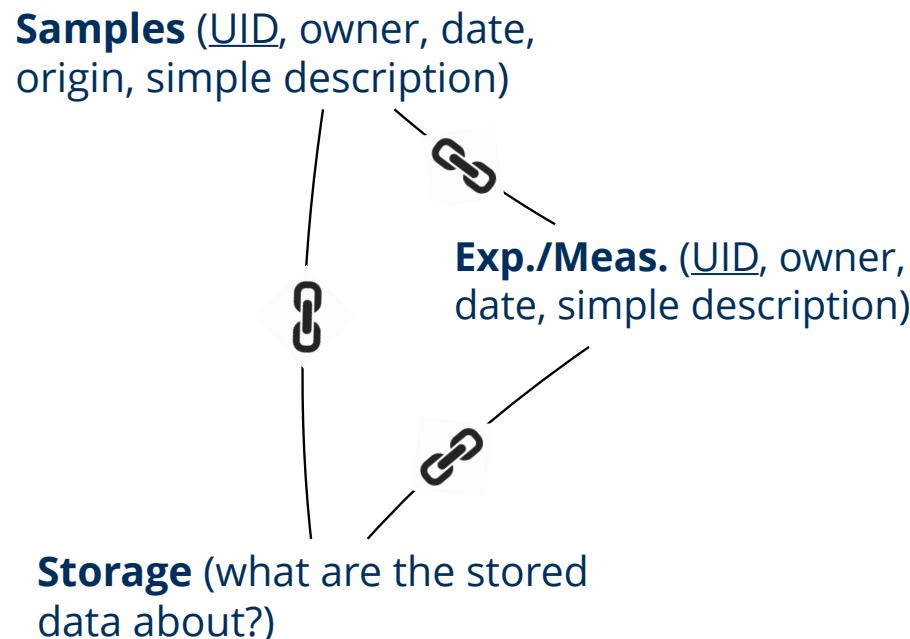
# To start with:



Maximal flexibility, in a simple,  
APAP and searchable way!!

# To start with:

**Minimal necessary** information:



# How to try to put some order with an e-lab tool: elabFTW



## Experiments

Expand all - Select all

Filter category Filter status Filter owner Filter visibility 30 Tags Show archived Go

Date Title Next step Category Status Tags Rating Owner

Everything that produces data

## Resources

Expand all - Select all

Filter category Filter status Filter owner Filter visibility 30 Tags Show archived Go

Date Title Next step Category Status Tags Rating Owner

2024-04-30  R785 embedded tactile hair 5\_I3 for AFM

SAMPLES AFM epofix tactile hair tarsus ventral Helen Rothfuss

Any matter object of a measurement

Minimal necessary information + Full documentation!!

# elabFTW — Samples

Minimal Info:

UID — 'R' + database ID

Simple description — Title

Date — Timestamp

Origin — Description

Owner

Bonuses:

Link — Parent/child sample and exp

Identifiable — QR !!



The screenshot shows the 'Editing an entry' page. Key fields include 'Started on' (12/05/2024), 'ID' (786), and 'Title' (Untitled). A context menu is open over the 'Title' field, showing options like PDF File, ZIP Archive, CSV File, and QR Code. Below the title, the 'Category' is set to 'Samples'. The 'Status' is 'Not set'. Under 'Tags', there is a button to 'Add a tag'. A rich text editor is present for the 'Sample description:' field, which contains the placeholder text 'Sample description:'. At the bottom, there are visibility and writing permissions settings.

The screenshot shows the 'Overview' page of the elabFTW database. It lists a single sample entry: "R785 embedded tactile hair 5\_1 for AFM". The entry includes details such as "Date: 2024-04-30", "Title: R785 embedded tactile hair 5\_1 for AFM", and "Owner: Helen Rothfuss". Other columns visible in the overview include "Next step", "Category", "Status", "Tags", "Rating", and "Owner". The "Owner" column is highlighted in purple. The "Tags" column for this entry lists "epofix", "tactile hair", "tarsus", and "ventral".

# elabFTW — identifiable samples!!



Started on 2024-04-12  
R760 Chitin dissolved in HFIP on Si wafers

Team: Default team  
Category: **SAMPLES**  
Status: NOT SET

Visibility: Only members of the team  
Can write: Only members of the team

**MAIN TEXT**

**UID:**  
R760

**Owner:**

**Purpose:**  
Samples were prepared to pass them on to **Yohaan Kudarkar** to perform ToFSIMS, to check whether the chitin is soluble in HFIP.

**Safety considerations:**

1. Wear all the protective equipment (lab coat, gloves, and goggles)
2. Work in the fumecupboard
3. When performing filtration, make sure that you are using fumecupboard in such a way that the glassware does not break.

**Sample description:**

**Preparations**

1. HFIP solution is precooled in an ice bath
  1. Cooling HFIP in an ice bath reduces its vapor pressure and it is much easier to handle e.g. when filtering.
  2. HFIP the same as used in [Samples - R664 Chitin in HFIP solution](#)
2. Chitin dissolved in HFIP is precooled in an ice bath
  1. [Samples - R664 Chitin in HFIP solution](#)
3. Whenever I am referring to eppendorf tubes, in this work it is always the one with a scrap cap

# elabFTW — experiments

Minimal Info:

UID — 'E' + database ID

Simple description — Title

Date — Timestamp

Owner

Experiments

Expand all - Select all

Date	Title	Next step	Category	Status	Tags	Rating	Owner
<input type="checkbox"/> 2024-05-23	<input type="button"/> E676-CF0-MN - ToF-SIMS feasibility test pad imprint			<span>SUCCESS</span>	<span>Cupiennius salei</span> <span>imprint</span> <span>Pad</span>		Carolin Fischer

Experiments > Editing an entry

Started on: 25/05/2024

ID: 677

Title: Untitled

Category: Not set

Status: Not set

Tags: Add a tag

Suggested tags:

Visibility: Only members of the team

Can write: Only owner and admins

MAIN TEXT

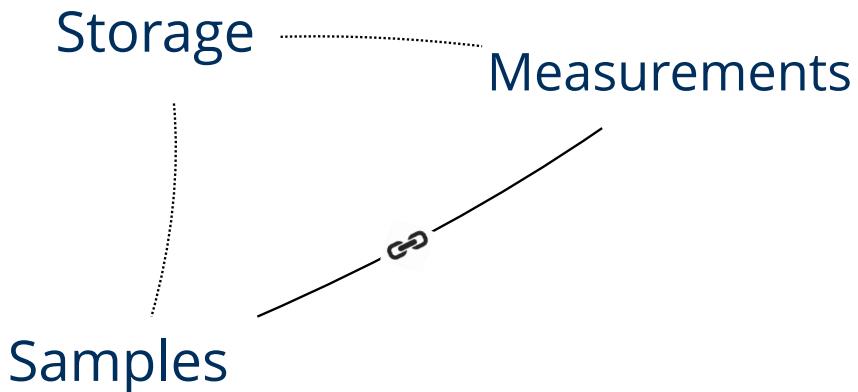
File Edit View Insert Format Tools Table

Goal :

Procedure :

Results :

# elabFTW: Linking samples and experiments



Started on 2024-04-18

## R770 TEM grids with ChNC negatively stained

### ▼ LINKED EXPERIMENTS

- ⌚ E589 Extraction of ChNC after dialysis - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3)
- ⌚ E620 Tip sonication and DLS on ChNC colloidal suspension - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3)

### ▼ LINKED RESOURCES

- ⌚ SP PROTOCOLS R108 2022-04-07-AuNa-TEM-grid-preparation

### ▼ RELATED EXPERIMENTS

- ⌚ EM E657 TEM observation of ChNCs at CRTD - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3)

[≡ Show related experiments](#)

Started on 2024-05-06

## E657 TEM observation of ChNCs

### ▼ LINKED EXPERIMENTS

- ⌚ E589 Extraction of ChNC after dialysis - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3)
- ⌚ E620 Tip sonication and DLS on ChNC colloidal suspension - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3)

### ▼ LINKED RESOURCES

- ⌚ SAMPLES R770 TEM grids with ChNC negatively stained prepared at CRTD

### Sample description:

Samples originate from E620 Tip sonication and DLS on ChNC colloidal suspension - Hydrolysis of Never Dried Non-deproteinised shrimp shells (H3); they undergo tip sonication.

Entry	Sample	Preparation to get 0.0005 wt%	Position in box no.	Staining date
R770-1	H3_4C_ChNC Ts	6.90ul at 0.1 wt% (10ul micropipette) in 700ul MQ (1000ul micropipette)	L6-6	Q8 Fri, 19 Apr 2024, 09:00

1.Followed the general procedure SP protocols - R108 2022-04-07-AuNa-TEM-grid-preparation

Entry	Sample	Preparation to get 0.0005 wt%	Position in box no.	Staining date
R770-2	H3_4C_ChNC Ts	same solution than R770-1	L6-14	A7 Thu, 2 May 2024, 09:00

Entry	Sample	Preparation to get 0.0005 wt%	Position in box no.	Staining date
R770-3	H3_1NC_ChNC	50ul at 0.003wt% (100ul micropipette) in 250ul MQ (1000ul micropipette)	L6-17	B4 Tue, 7 May 2024, 09:00
R770-4	H3_1C_ChNC	50ul at 0.003wt% (100ul micropipette) in 250ul MQ (1000ul micropipette)	L6-18	B6 Tue, 7 May 2024, 09:00

Comment: R770-3 fell on the coated side on the box when trying to put it inside.

tinetti

27.05.2024

# Adding metadata to the data — python API

## Raw data storage folder structure

```
User Name 1
  - Experiments
    + Technique1
      # Exyy
    + Technique 2
      # Exyz
    ...
  ...
  - Samples
    # Rxyy
    # Rxyz
  ...
User Name 2
```

- Read the users from the DB
- Read the experiments of a user
- Read the folders of that user
- For each experiment look if a folder containing the exp UID can be found — If yes, export the pdf

# Adding metadata to the data – python API

```
raw_data_folder = r'\\cmcb-files\\bcube_politi_rawdata'
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
dry_run = False
for un,ufn in zip(lofusr_fn,lofusr_full):
    for fp in filelist:
        if (un == fp or ufn == fp) and ('\\fn' not in un):
            print('processing directory of user: '+ufn)
            n_path = os.path.join(raw_data_folder,fp)
            if not os.path.isdir(n_path):
                os.makedirs(n_path)
            if not os.path.isdir(os.path.join(n_path)):
                os.makedirs(os.path.join(n_path))
            # s path = os.path.join(n_path,'Experiments')
            # for x in os.walk(n_path):
            #     if len(x[1])>0 and len(x[1])<count(os.sep):
            #         print('problem for: '+x[0])
            #         if exp.fullname == ufn:
            #             osr_lofres_dbn = [res for res in itemlist if res.fullname == ufn]
            #             for exp in osr_lofres_dbn:
            #                 if len(exp) >=1:
            #                     idxs = [idx for idx,s in enumerate(exp.lof_f) if str('E'+str(exp.id)) in s.split(os.sep)[-1]]
            #                     if len(idxs) >=1:
            #                         out_path = exp.lof_f[idxs[0]]
            #                     if len(idxs)>1:
            #                         l_of_dup_levs = [usr_lof_f[idx][len(n_path)],count(os.sep) for idx in idxs]
            #                         if all(x == 1 of dup_levs[0] for x in l_of_dup_levs):
            #                             print('problem for: '+exp.id,exp.fullname,idxs)
            #                         else:
            #                             print('found experiment: ',exp.id,exp.fullname,out_path)
            #                             filename = f'E{exp.id}-(exp.modified_at)-export.pdf'.replace(':','')
            #                             if filename not in os.listdir(out_path):
            #                                 export_item(experimentsApi,exp,out_path,filename,dryrun=dry_run)
            #             elif len(idxs)==1:
            #                 print('found experiment: ',exp.id,exp.fullname,out_path)
            #                 filename = f'{exp.id}-(exp.modified_at)-export.pdf'.replace(':','')
            #                 if filename not in os.listdir(out_path):
            #                     export_item(experimentsApi,exp,out_path,filename,dryrun=dry_run)
            for res in usr_lofres_dbn:
                idxs = [idx for idx,s in enumerate(res.lof_f) if str('R'+str(res.id)) in s.split(os.sep)[-1]]
                if len(idxs) >=1:
                    out_path = res.lof_f[idxs[0]]
                if len(idxs)>1:
                    l_of_dup_levs = [usr_lof_f[idx][len(n_path)],count(os.sep) for idx in idxs]
                    if all(x == 1 of dup_levs[0] for x in l_of_dup_levs):
                        print('problem for: ',res.id,res.fullname,idxs)
                    else:
                        print('found resource: ',res.id,res.fullname,out_path)
                        filename = f'R{res.id}-(res.modified_at)-export.pdf'.replace(':','')
                        if filename not in os.listdir(out_path):
                            export_item(items,res,out_path,filename,dryrun=dry_run)
            elif len(idxs)==1:
                print('found resource: ',res.id,res.fullname,out_path)
                filename = f'{res.id}-(res.modified_at)-export.pdf'.replace(':','')
                if filename not in os.listdir(out_path):
                    export_item(items,res,out_path,filename,dryrun=dry_run)
```



Name	Date modified	Type	Size
E554-01	2/29/2024 10:54 AM	File folder	
E554-02	2/29/2024 10:54 AM	File folder	
E554-2024-03-26 171802 export.pdf	5/10/2024 5:37 PM	Foxit PDF Reader ...	38 K

E554 DVM6 of chitin from HFP solution

Date: 2024-02-28

- To double check film forming properties of HFP sample

## Results

- The solutions definitely have larger chunks of stuff which is not dissolved well
  - The black dots etc probably come from the dirty glass slide even though I cleaned them a bit with ethanol

## Next steps

- Centrifuge samples properly
  - When performing experiments to dry films on Si wafers, clean the surfaces extremely well with ethanol/water, potentially even detergent.

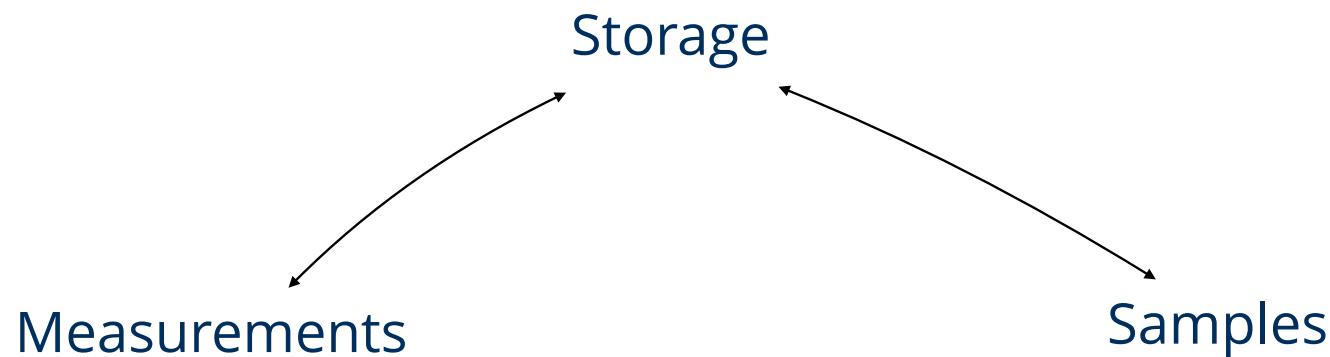
### **Procedure**

E554-6

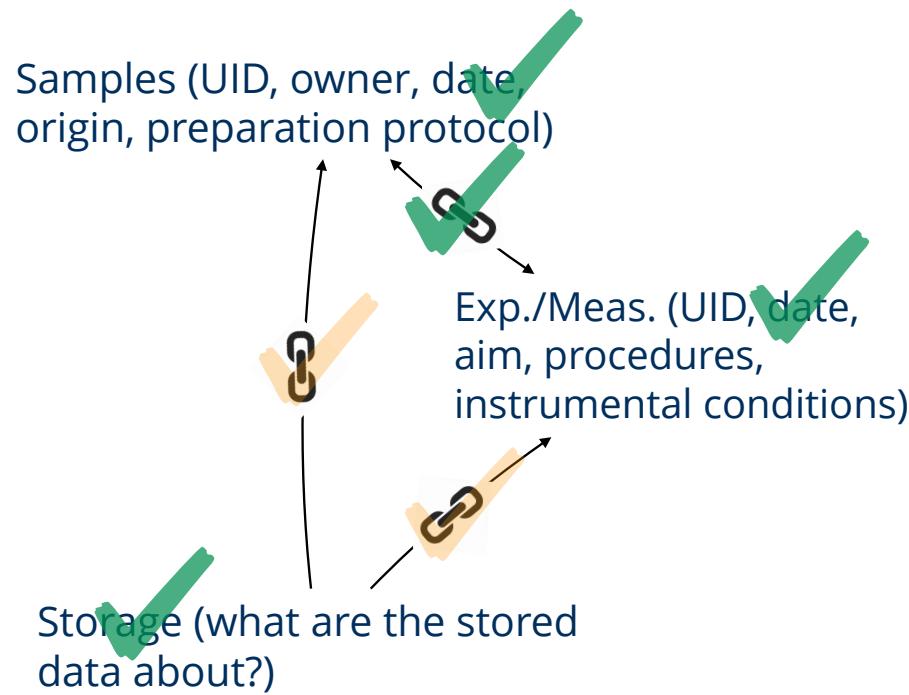
1. Looking at s582 sample
    1. Centrifuged and 50  $\mu$ L supernatant placed on a glass slice and dried in a closed glass petri dish at RT in a fume cupboard.

# Adding the data storage location to the documentation — python API

TB implemented in  
the next future



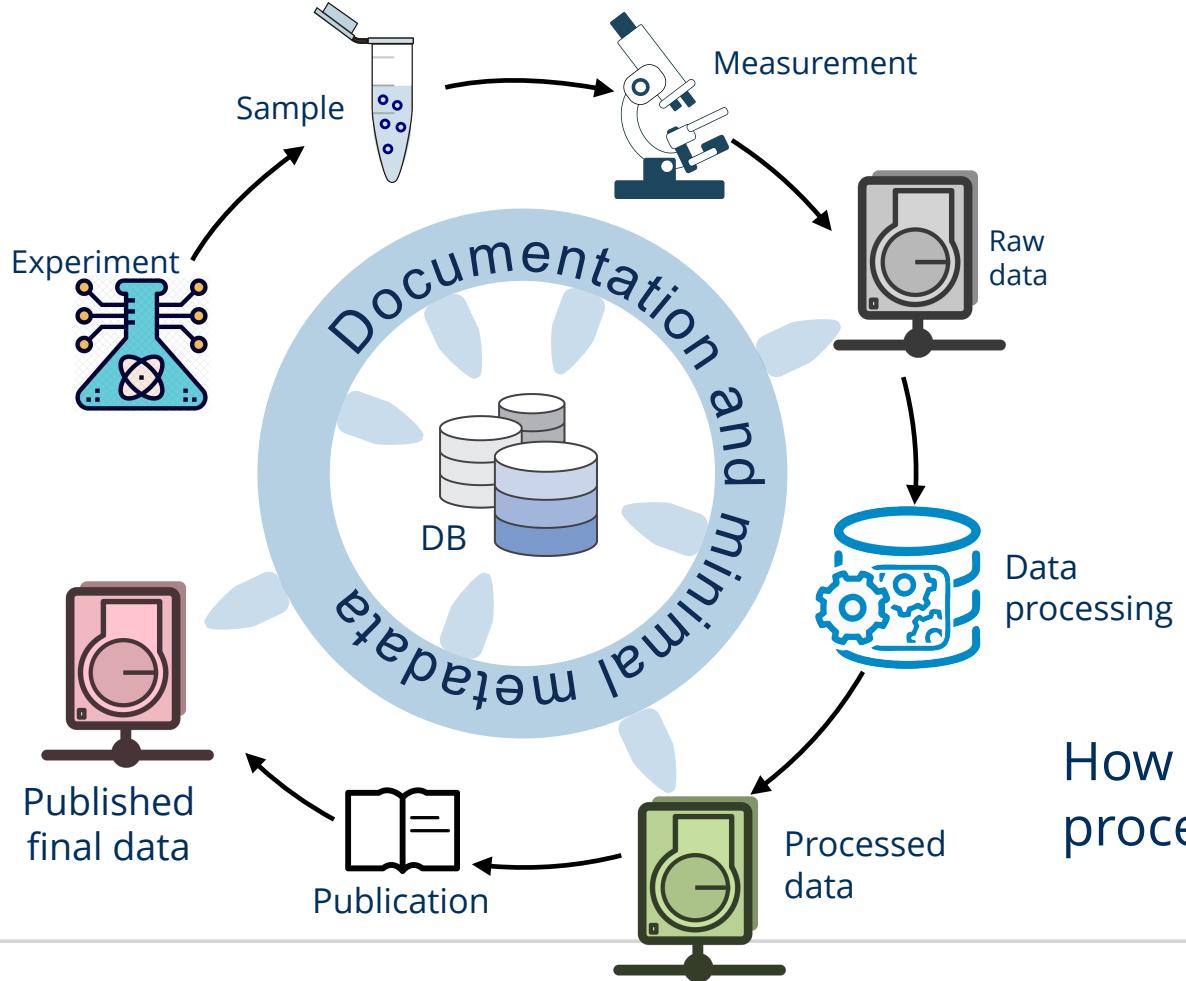
# Almost there!!



But still far behind in:

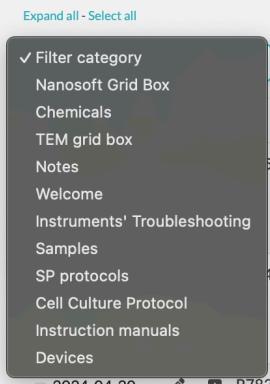
- Storing standard metadata

# Still a long way to go



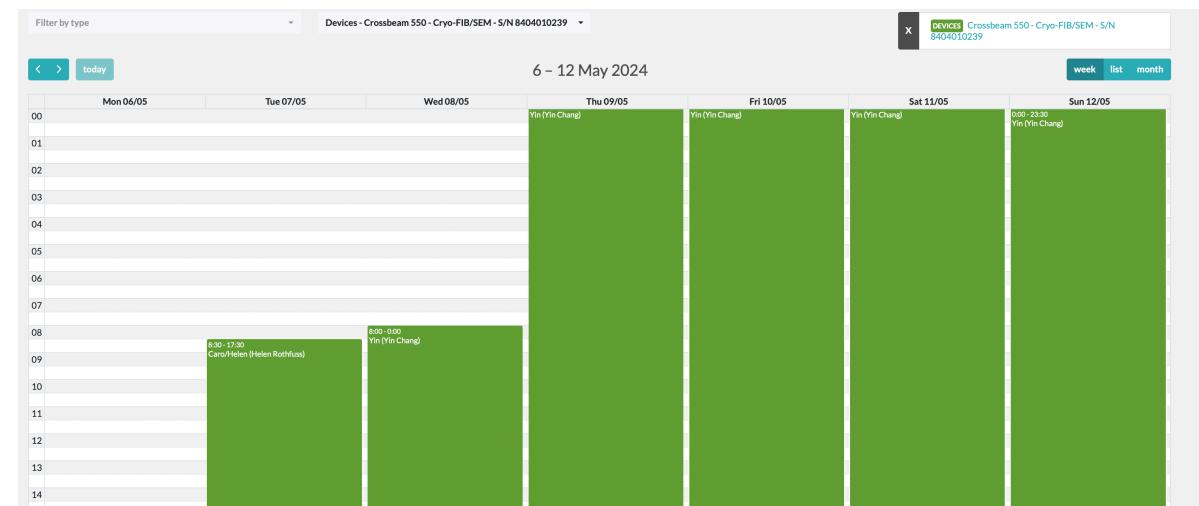
How to deal with  
processing code

## Resources



# On top of that:

- Full documentation
- Tagging
- Device management (booking, instruction manuals, troubleshooting)
- Sample storage management (what sample is where)
- Sample preparation protocols
- On boarding and off boarding



# Advantages and challenges

Very effective and flexible tool to document scientific activity (readability and searchability)

Easy 2-way link between experiment and resources

Simple API can be used to:

- store the metadata with the data and
- add the location of the data on the documentation

Great tool to:

- manage devices, instruction manuals and sample preparation protocols etc...
- On boarding and off boarding of new/leaving members

- Takes some time to find an effective way of documenting/categorising/tagging items
- Not always straightforward to add schemes, pictures or printouts to the documentation
- Sometimes, breaking changes with the upgrade
- Tools still under heavy development
- Requires some programming skills to use the API

# Acknowledgments

Aurimas Narkevicius



Politi Group @ BCUBE

Alice Ludewig



CMCB IT team



Thank you!



BCUBE<sup>3</sup>  
Center for Molecular  
Bioengineering TU Dresden