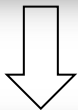


Documentation and Metadata

Why would I use an ELN for documentation?



Paper replacement

Project-based system

Share data

Experiments, protocols, ...

Standardization of experiments

Templates

Intellectual property protection

Data stored in central database

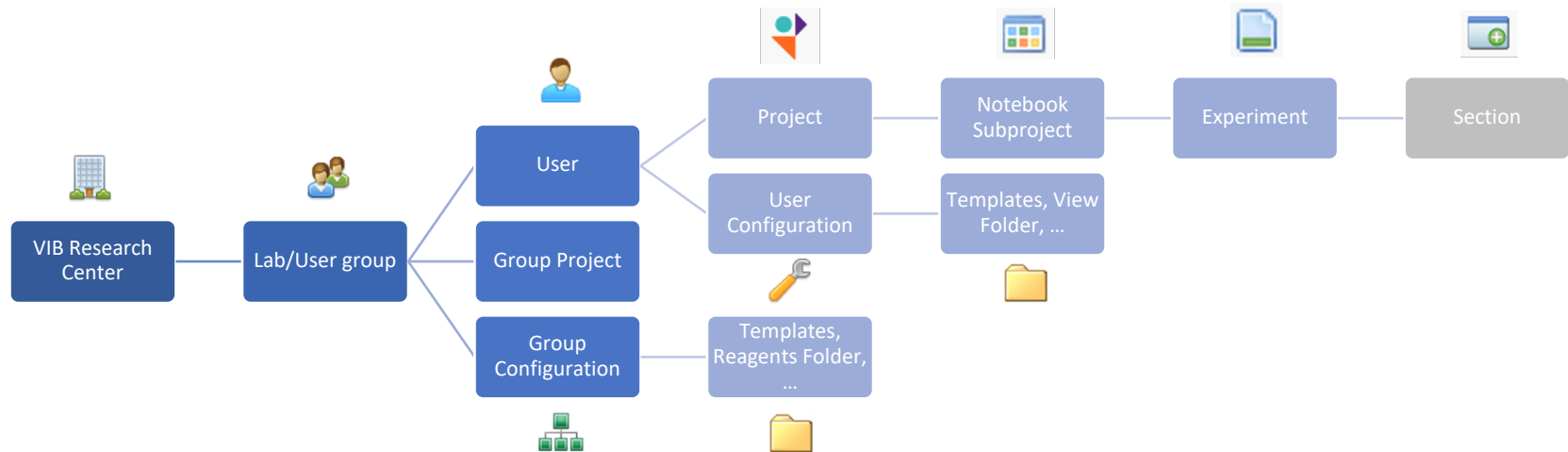
Eventually Office integration

MS Office: Word, Excel, PowerPoint

Search function

Advanced search

Project-based system



General consideration with the publishing in mind

- Write the **documentation** in such a way that someone else who is known to the field can **not mis-interpret** any of the data, even if they tried.
- Documentation at two levels
 - Project/Study level:
 - title, summary, aims, authors, funds
 - methods, license and id for “data sets”, folder structure, file naming
 - Data-level documentation

General consideration with the publishing in mind

- Both the study- and data-level documentation must be generated as early as possible in the research process and also maintained, in order to be accurate and complete
- Further reading:
https://rdmkit.elixir-europe.org/metadata_management



DMP

Contact

DMP Staff



"Solution" with fields

DMP

Contact

DMP Staff



How to write useful documentation?

- Lab protocols in an example publication

The resulting emulsion was collected in aliquots of 50 μ L total volume and thermocycled according to the RT program (42°C for 90 min, 11 cycles of [50°C for 2 min, 42°C for 2 min], 85°C for 5 min, followed by a final hold on 4°C). 125 μ L of recovery agent (20% PFO in HFE), 55 μ L of GITC Buffer (5 M GITC, 25 mM EDTA, 50 mM Tris-HCl pH 7.4) and 5 μ L of 1 M DTT was added to each separate aliquot of 50 μ L thermocycled emulsion and incubated on ice for 5 min.

How to write useful documentation?

- Lab protocols in an example publication

The resulting emulsion was collected in aliquots of 50 μ L total volume and thermocycled according to the RT program (42°C for 90 min, 11 cycles of [50°C for 2 min, 42°C for 2 min], 85°C for 5 min, followed by a final hold on 4°C). 125 μ L of recovery agent (20% PFO in HFE), 55 μ L of GITC Buffer (5 M GITC, 25 mM EDTA, 50 mM Tris-HCl pH 7.4) and 5 μ L of 1 M DTT was added to each separate aliquot of 50 μ L thermocycled emulsion and incubated on ice for 5 min.

see here: <https://www.protocols.io/view/hydrop-rna-v1-o-dm6gpwqjilzp/v2?step=5>

Extraction of DNA from soil samples

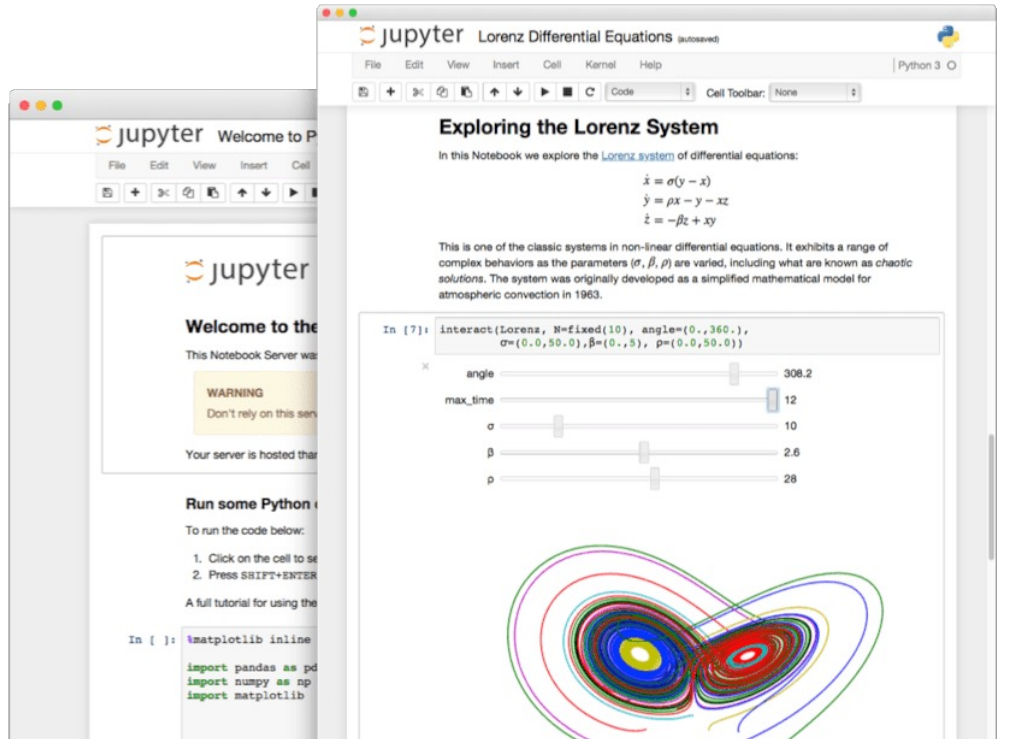
<https://www.protocols.io/view/soil-sample-dna-extraction-supersoil-rm7vzbzj2vx1/v1>

How to document scripts and code

- Data analysis scripts

Barcode reads were trimmed to exclude the intersub-barcode linear amplification adapters using a mawk script. Reads were then mapped and cell-demultiplexed using STARsolo ([Kaminow et al., 2021](#)) in CB_UMI_Complex mode. The resulting STARsolo-filtered count matrices were further analyzed using Scanpy ([Wolf et al., 2018](#)). In short, cells were filtered on expression of a maximum of 4000 genes, and a maximum of 1% UMIs from mitochondrial genes. Genes were filtered on expression in a minimum of three cells. Potential cell doublets were filtered out using a Scrublet ([Wolock et al., 2019](#)) threshold of 0.25.

Jupyter Notebooks: very eloquent

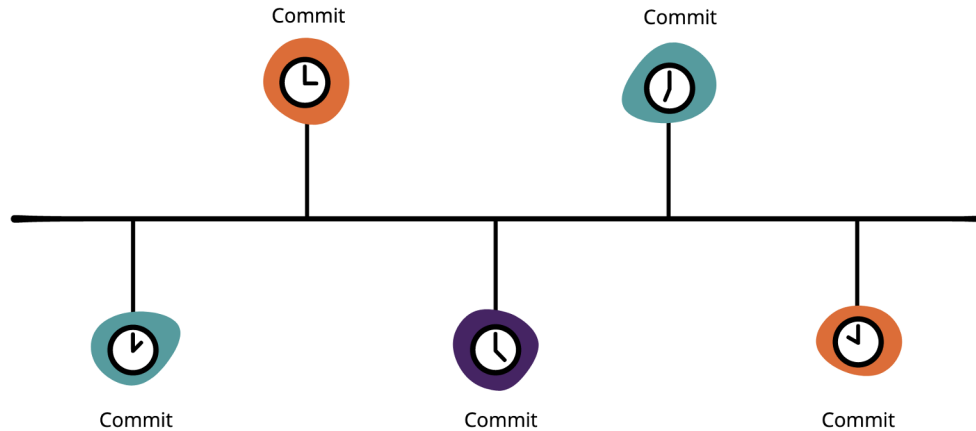


Data analysis scripts under version control

The best possible way of documenting scripts and code for software is using git and github (or alternative online 'backup' solution to github).



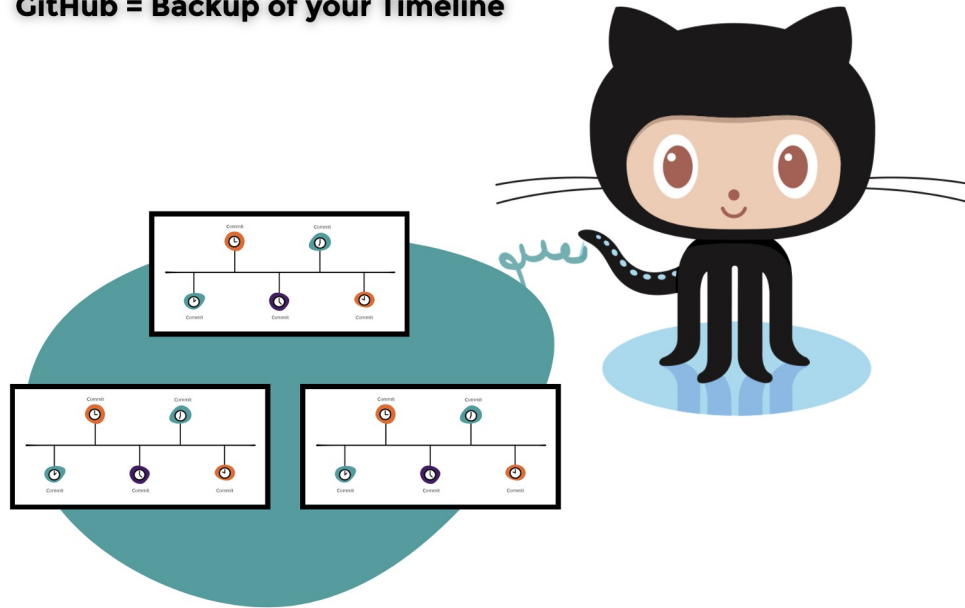
Git Repository = Timeline



Data analysis scripts under version control

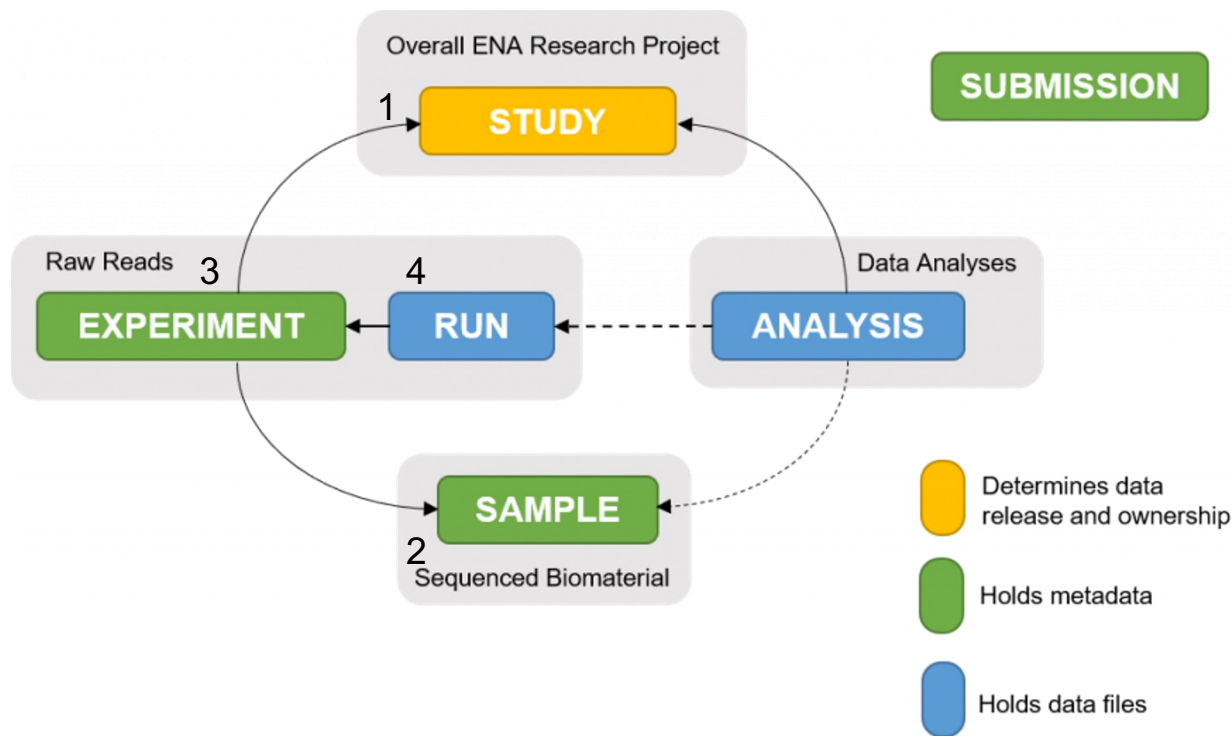
The best possible way of documenting scripts and code for software is using git and github (or alternative online backup solution to github).

GitHub = Backup of your Timeline

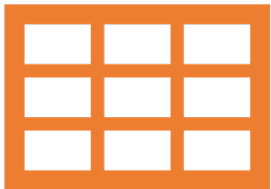


What do you know about metadata?

ENA's metadata model as an example



Let's focus on the metadata checklists



- [From ENA](#)
- [ELIXIR's Fair Cookbook for transcriptomics data](#)

Documentation and support

VIB & ELIXIR Support

training@vib.be

Research Data Service Facility has been set up in 2023

Contact details of RDM support of the Flemish universities:

<https://elearning.bits.vib.be/courses/writing-a-data-management-plan/lessons/rdm-support/topic/contact-details/>

ELIXIR services - usegalaxy.eu

More information at: <https://www.elixir-belgium.org/services>

What did we cover today.



implement SOP type of approach for your daily documentation of experiments



describe the impact of documentation on the publication preparation



make versioning more persistent by using protocols.io and/or your Electronic Lab Notebook



use github for scripts and code

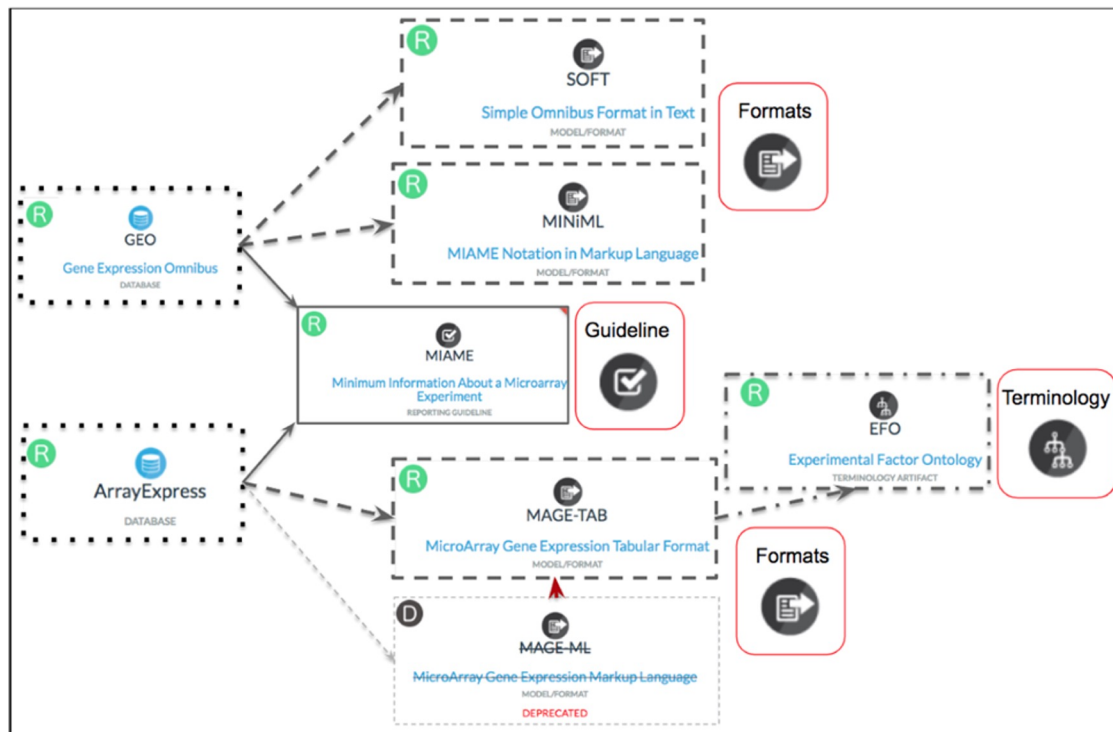


apply at least minimal metadata standards for domain-specific data

External references

- Harvard Comparison Grid:
<https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>
- NKI: LabGuru for Wetlab and Castor for Clinical Data Capture
- MDC and BIH – LIMS is available, RDM SODAR (Omics data)

First a short excursion to nomenclature



Metadata
model

Metadata
checklist

There are many metadata models

