



Helis Academy

Organising and documenting data with an ELN

Hanne Vlietinck



UHASSELT

KNOWLEDGE IN ACTION



Subjects

- ✓ **Organising and documenting data (basics)**
- ✓ **ELN (basics)**
(Hanne Vlietinck, Data Steward, U Hasselt)
- ✓ **Demo Benchling**
(Jonas Delva, PhD Student Life Sciences, UGent)
- ✓ **Demo ELabFTW**
(Ing. Nicolas Carpi, Engineer and Founder ELabFTW, Institute Curie Paris)
- ✓ **Experience of a PhD student (ELabFTW)**
(Ing. Robbe Breugelmans, PhD student U Hasselt)



Learning outcomes

- *Importance and good practices organized, documented data*
- *Basics of ELNs and how to choose an ELN*
- *Basics of Benchling, ELabFTW*

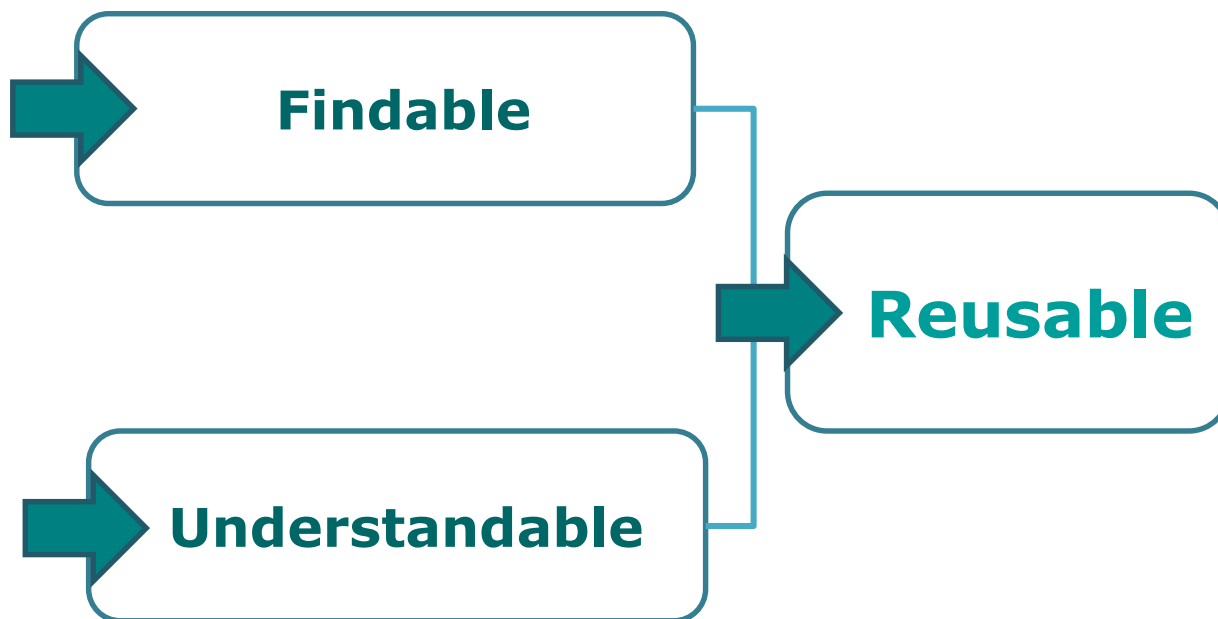


Organising, documenting data with an ELN

Importance of organised/ documented data



***Prevent
Data loss!!***





Organising, documenting data with an ELN

A FAVOUR, TO YOUR FUTURE SELF

*Do yourself, your future self and society a favour
by organising and documenting your data*

Your future self will be thankful to you





Organising, documenting data with an ELN

Good practices

Organising

- ✓ Folders
- ✓ File naming
- ✓ File versioning

**Easily searchable
Findable**

Documenting

- ✓ Readme-files
- ✓ Other data documentation
- ✓ Automated Metadata

Understandable

Reusable



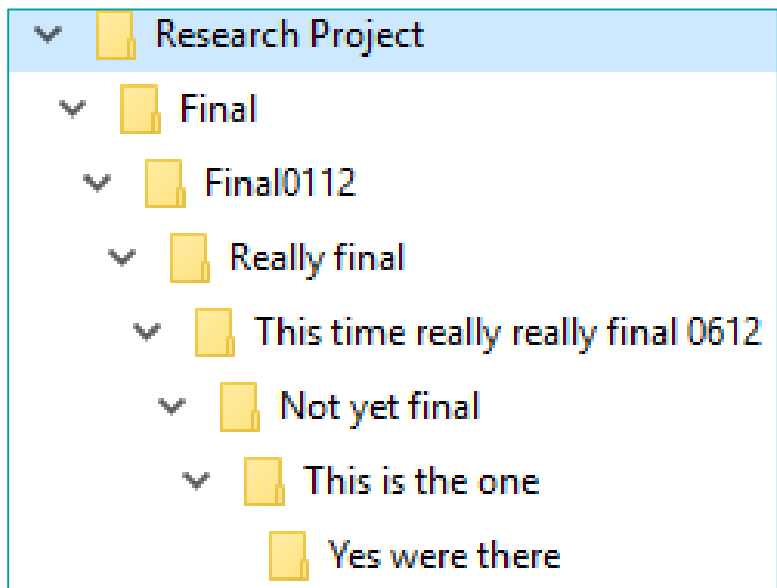


Organising data: Folder structure



Bad practice...

NOT Findable!
NOT Reusable!



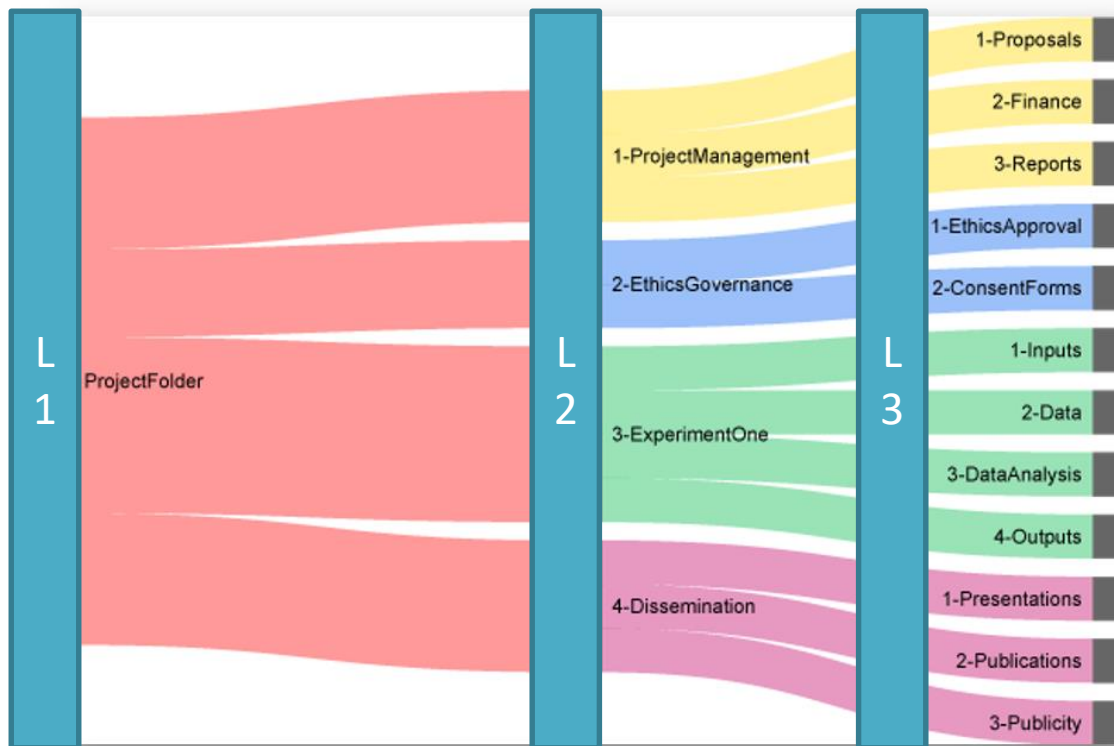


Organising, documenting data with an ELN

Organising data: Folder structure

Good practice

Original slide by Marta Teperek



Max 3 levels

Copyright: <http://www.vukovicnikola.info/folder-structure-for-research/>



Organising, documenting data with an ELN

Organising data: File naming convention



Good practice

Name	Example
Month_year_type	01_1989_landsat
Month_year_type_process	01_1989_landsat_orthorect
Field_points_date of actualization	Field_points_11_09_13
Month_year_outcome type	01_1989_wateryield
Year_parameter	2050_wateryield
Parameter_date of actualization	Rainfall_11_09_2013
Model_script_version	Shetran_script_v2.5
Model_script_modification_trial#	Shetran_script_cloudwater_t1
Parameter_col#_row#_trial#	Rainfall_4_250_t1
Est. parameter_col#_row#_trial#	ET_4_250_t1

Source: Wageningen University and Research



Organising data: File versioning

Possible applications





Organising, documenting data with an ELN

Good practices

Organising

- ✓ Folders
- ✓ File naming
- ✓ File versioning

**Easy searchable
Findable**

Documenting

- ✓ Readme-files
- ✓ Other data documentation
- ✓ Automated Metadata

Understandable

Reusable





Documenting the 5 W's en 2 H's

- ✓ What
- ✓ When
- ✓ Where
- ✓ Who
- ✓ Why
- ✓ How
- ✓ How much





Documenting data

Data documentation

General Readme-files

Other data documentation

- ✓ Folder structure
- ✓ File naming convention
- ✓ Methodology
- ✓ Protocols, Variables
- ✓ Index (Abbreviations)
- ✓ Ontologies

Automated Metadata





Documenting data: Readme files (General)

Best practice

- ✓ At least 1 readme per dataset
- ✓ In the beginning
& updates
during the research life cycle





Documenting data: Readme files (General)

Harvard Template

1. **Dataset title**
2. **Name/institution/contact information (PI)**
3. **File name structure:**
 - **Provide the template**
 - **Attributes:** Describe the attributes used to name the files.
 - **Codes:** Provide a complete list of any codes/abbreviations used.
 - **Provide examples of above items.**
4. **File formats**
5. **Column headings for Tabular data** list & define:
 - **Units of measurement**
 - **Data formats** ex. YYYYMMDD
 - **Calculations**
6. **Versioning**

Source: Harvard <https://datamanagement.hms.harvard.edu/readme-files>





Documenting data: Readme files (General)

Example

- 1. Dataset title:** Raw Images for Experiment A, Smith
- 2. Principal Investigator:** John Smith, PI, 555-555-5555
, jsmith@hms.harvard.edu
- 3. File name structure**
ExperimentName_InstrumentID_CaptureDateTime_ImageID.tif
The base file name is composed of the name of the experiment, the ID number of the instrument used, the date and time that the image was captured, and the unique identifier of the image.

Attributes:
ExperimentName = Name of the experiment.
Instrument ID = Five-digit code assigned to the lab instrument.
See the Codes section for a list of instruments and their ID numbers.
CaptureDateTime = Date and time at which the image was captured, in YYYYMMDDThhmm format.
Image ID = Three-digit unique identifier for image, such as 001, 002,..
Codes:
[List of instruments and IDs]
Examples:
daf2-age1_14052_20150412T0515_005.tif
- 4. File formats:** .tiff

Source: Harvard <https://datamanagement.hms.harvard.edu/readme-files>





Organising, documenting data with an ELN

Documenting data: Folder structure

```
├── README.md           : Description of this repository
├── LICENSE             : Repository license
├── standard            : All content considered part of the Darwin Core standard
│   ├── vocabularies
│   │   └── term_versions.csv : Darwin Core term versions, contains the normative definitions of the terms
│   └── documents
│       ├── simple        : Simple Darwin Core guide
│       ├── rdf           : RDF guide
│       ├── text          : Text guide
│       └── xml           : XML guide
├── www                : GENERATED Darwin Core website
│   └── guides
│       └── index.html    : GENERATED Darwin Core quick reference guide
├── dist               : GENERATED Distribution files generated by build.py
│   ├── dwc_terms.csv    : GENERATED CSV file with the recommended Darwin Core terms, definitions, etc.
│   ├── simple_dwc_horizontal.csv : GENERATED CSV file with Simple Darwin Core terms as a row
│   └── simple_dwc_vertical.csv : GENERATED CSV file with Simple Darwin Core terms as a column
├── build
│   ├── build.py         : Build script to generate distribution files from the normative document
│   └── config
│       ├── index.tmpl   : Template for index.html
│       └── terms.csv     : Configuration for the recommended terms, incl. order, labels, examples
├── .github
│   ├── CONTRIBUTING.md  : Guide on how to contribute to Darwin Core
│   └── ISSUE_TEMPLATE.md : Template for issues
└── .gitignore          : Files and directories to be ignored by git
```



Organising, documenting data with an ELN

Documenting data: File naming convention

Name	Example
Month_year_type	01_1989_landsat
Month_year_type_process	01_1989_landsat_orthorect
Field_points_date of actualization	Field_points_11_09_13
Month_year_outcome type	01_1989_wateryield
Year_parameter	2050_wateryield
Parameter_date of actualization	Rainfall_11_09_2013
Model_script_version	Shetran_script_v2.5
Model_script_modification_trial#	Shetran_script_cloudwater_t1
Parameter_col#_row#_trial#	Rainfall_4_250_t1
Est. parameter_col#_row#_trial#	ET_4_250_t1

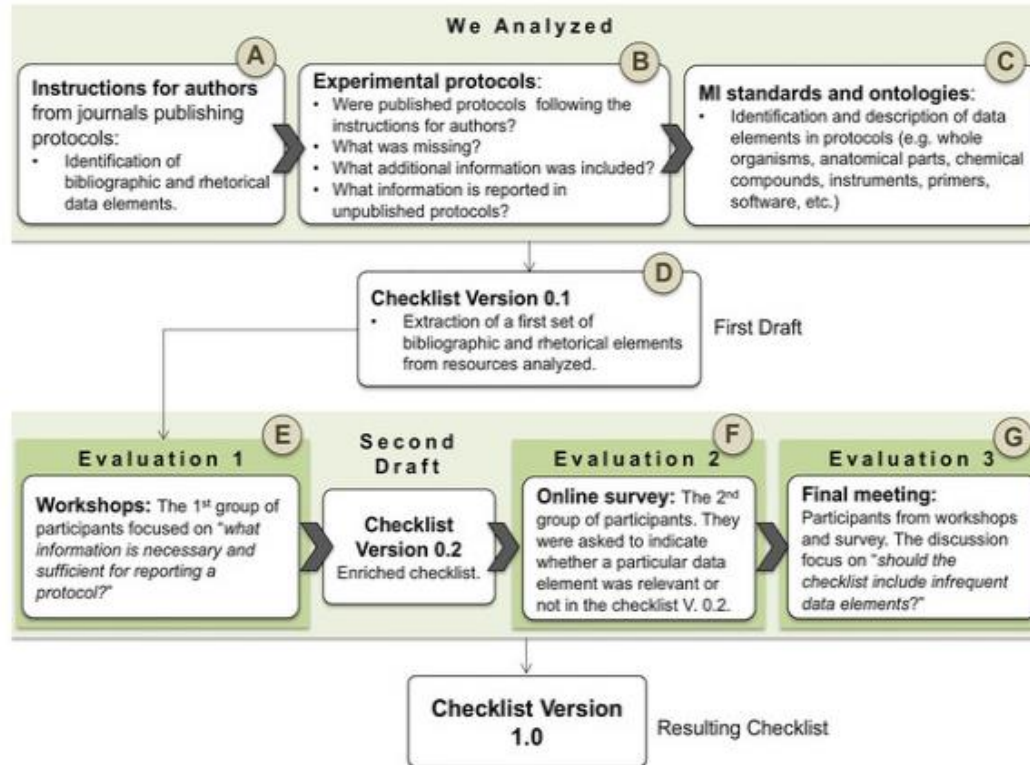
Source: Wageningen University and Research





Organising, documenting data with an ELN

Documenting data: Methodology



Source: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com) open access, cc



Organising, documenting data with an ELN

Documenting data: Protocols

Data element	Property
Title of the protocol	
Author	Name Identifier
Version number	
License of the protocol	
Provenance of the protocol	
Overall objective or purpose	
Application of the protocol	
Advantage(s) of the protocol	
Limitation(s) of the protocol	
Organism	Whole organism / Organism part Sample/organism identifier Strain, genotype or line Amount of Bio-Source Developmental stage Bio-source supplier Growth substrates Growth environment Growth time Sample pre-treatment or sample preparation

Source: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com*) open access, cc



Organising, documenting data with an ELN

Documenting data: Protocols

Data element	Property
Laboratory equipment	Name
	Manufacturer or vendor (including homepage)
	Identifier (catalog number or model)
	Equipment configuration
Laboratory consumable	Name
	Manufacturer or vendor (including homepage)
	Identifier (catalog number)
Reagent	Name
	Manufacturer or vendor (including homepage)
	Identifier (catalog number)
Kit	Name
	Manufacturer or vendor (including homepage)
	Identifier (catalog number)
Recipe for solution	Name
	Reagent or chemical compound name
	Initial concentration of a chemical compound
	Final concentration of chemical compound
	Storage conditions
	Cautions
	Hints

Source: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com) open access, cc



Organising, documenting data with an ELN

Documenting data: Protocols

Data element	Property
Software	Name Version number Homepage
Procedure	List of steps in numerical order Alternative/Optional/Parallel steps Critical steps Pause point Timing Hints Troubleshooting

Source: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com) open access, cc



Documenting data: Index/ Register

Excel file

B	SoP	Can-f-1	Fel-d-1	So	Ca	Ma
<u>Abessijn</u>	c	null	1	135	100	70
<u>American Curl</u>	c	0	1	135	100	70
<u>Amerikaans draadhaar, ook wel American wirehair</u>	C	0	1	135	100	70
<u>Amerikaans korthaar, ook wel American shorthair</u>	c	0	1	135	100	70
<u>Amerikaanse stompstaartkat, ook wel American bobtail</u>	c	0	1	135	100	70
<u>Asian</u>	c	0	1	135	100	70
<u>Balinees (langharige siamees)</u>	c	null	1	135	100	70
<u>Bengaal</u>	c	0	1	135	100	70
<u>Blauwe Rus</u>	C		1	135	100	70
<u>Bombay</u>	c		1	135	100	70
<u>Brits korthaar</u>	c	0	1	135	100	70
<u>Burmees</u>	c	/	1	120	90	60
<u>Cornish rex</u>	c	0	1	120	90	60
<u>Chartreux, ook wel karthuizer</u>	c	0	1	120	90	60
<u>Devon Rex</u>	c	0	1	120	90	60
<u>Don sphynx</u>	c	/	1	120	90	60
<u>Eqyptische mau</u>	c	0	1	120	90	60

No index



Not recommended...





Organising, documenting data with an ELN

Documenting data: Index/ Register

Excel file

No index



Not recommended...

B	SoP	Can-f-1	Fel-d-1	So	Ca	Ma
Abessijn	c	null	1	135	100	70
American Curl	c	0	1	135	100	70
Amerikaans draadhaar_ ook wel American wirehair	C	0	1	135	100	70
Amerikaans korthaar_ ook wel American shorthair	c	0	1	135	100	70
Amerikaanse stompsstaartkat_ ook wel American bobtail	c	0	1	135	100	70
Asian	c	0	1	135	100	70
Balinees (langharige siamees)	c	null	1	135	100	70
Bengaal	c	0	1	135	100	70
Blauwe Rus	C		1	135	100	70
Bombay	c		1	135	100	70
Brits korthaar	c	0	1	135	100	70
Burmees	c	/	1	120	90	60
Cornish rex	c	0	1	120	90	60
Chartreux_ ook wel karthuizer	c	0	1	120	90	60
Devon Rex	c	0	1	120	90	60
Don sphyinx	c	/	1	120	90	60
Egyptische mau	c	0	1	120	90	60



With index



Good practice

Index abbreviations

B	Breed
Sop	Sort of pet
Can-f-1	Canine allergy values
Fel-d-1	Feline allergy values
So	Sofa
Ca	Carpet
Ma	Matress





Organising, documenting data with an ELN

Documenting data: Ontologies

Ontology	Description
The Ontology for Biomedical Investigations (OBI) (Bandrowski et al., 2016)	An ontology for the description of life-science and clinical investigations.
The Information Artifact Ontology (IAO) (IAO, 2017)	An ontology of information entities.
The ontology of experiments (EXPO) (Soldatova & King, 2006)	An ontology about scientific experiments.
The ontology of experimental actions (EXACT)	An ontology representing experimental actions.
The BioAssay Ontology (BAO) (Abeyruwan et al., 2014)	An ontology describing biological assays.
The Experimental Factor Ontology (EFO) (Malone et al., 2010)	The ontology includes aspects of disease, anatomy, cell type, cell lines, chemical compounds and assay information.
eagle-i resource ontology (ERO)	An ontology of research resources such as instruments, protocols, reagents, animal models and biospecimens.
NCBI taxonomy (NCBITaxon) (Federhen, 2015)	An ontology representation of the NCBI organismal taxonomy.
Chemical Entities of Biological Interest (ChEBI) (Hastings et al., 2013)	Classification of molecular entities of biological interest focusing on 'small' chemical compounds.
Uberon multi-species anatomy ontology (UBERON) (Mungall et al., 2012)	A cross-species anatomy ontology covering animals and bridging multiple species-specific ontologies.
Cell Line Ontology (CLO) (Sarntivijai et al., 2014 ; Sarntivijai et al., 2011)	The ontology was developed to standardize and integrate cell line information.

Source: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com) open access, cc



Documenting data: Automated Metadata

DATA DOCUMENTATION

- ✓ Readme file
- ✓ Folder structure
- ✓ File naming convention
- ✓ Methodology
- ✓ Protocols, Variables
- ✓ Index (Abbreviations)
- ✓ Ontologies

Human readable

AUTOMATED METADATA

Standards

For example Dublin Core

Info like Creator, Keywords, Title, orchid, Grantnr,...

Machine readable
.xml, .json





Organising, documenting data with an ELN

Good practices

Organising

- ✓ Folders
- ✓ File naming
- ✓ File versioning

Documenting

- ✓ Readme-files
- ✓ Other data documentation
- ✓ Automated Metadata

ELN
Electronic Lab Notebook





What is an ELN?

ELN stands for Electronic lab notebook

- ✓ = **Computer program**
- ✓ **Replaces paper laboratory notebooks.**
- ✓ **To document organize research, experiments, procedures professionally**





Organising, documenting data with an ELN

Why should you use an ELN?

- ✓ *Easier to publish*
- ✓ *Efficient functionalities*

It makes researchers life easier





Organising, documenting data with an ELN

Why should you use an ELN?

➔ **Structured place**
(Search function, version control)

➔ **Easy Collaboration**



PLATFORM





Organising, documenting data with an ELN

Why should you use an ELN?



Prevent data loss



*Papers can be lost, damaged
Not easily readable by others
Not easy searchable*



Prevent unauthorized access



SECURITY



UHASSELT

KNOWLEDGE IN ACTION



Organising, documenting data with an ELN

Why should you use an ELN?



Extra Efficient functions

Calculations, Connection instruments templates



Automatic metadata creation



Timestamps

Legal document: court of law as evidence, patent

TIME SAVING



UHASSELT

KNOWLEDGE IN ACTION



Return of investment

*“Difficult to change old habits BUT
when ELN becomes common practice*

*Long-term benefits **outweigh** initial work investment.
Return of investment”*





Organising, documenting data with an ELN

Wide range of products

Cambridge/Harvard MATRIX

Analysed by:

- Cambridge
- Harvard
- Cambridge & Harvard

Benchling	Hivebench	Open Science Framework
Biovia	IDBS	OpenLab
BrightLab	LabArchives	OpenWetWare
Chemotion	LabCloud	OSF
Confluence	LabCollector	PerkingElmer
Docollab	Labfolder	Pillar Science
eclabnote	LabGuru	Quiver
e-Notebook	Labii	REDCap
e-Workbook	Labstep	Riffyn
eLabFTW	LabTrove	RSpace
eLABJournal	Labvantage	Scilligence
ELOG	Labware	SciNote
Evernote	Mbook	Signals
Exemplar	OneNote	SLIMS
Findings	Openbis	Studies Notebook





Organising, documenting data with an ELN

Wide range of products

- How to choose??-





How to choose an ELN?

JISC & GLASGOW model

MUST HAVES

SHOULD HAVES

COULD HAVES

General functionality

IT functionality

Collaboration

Integration





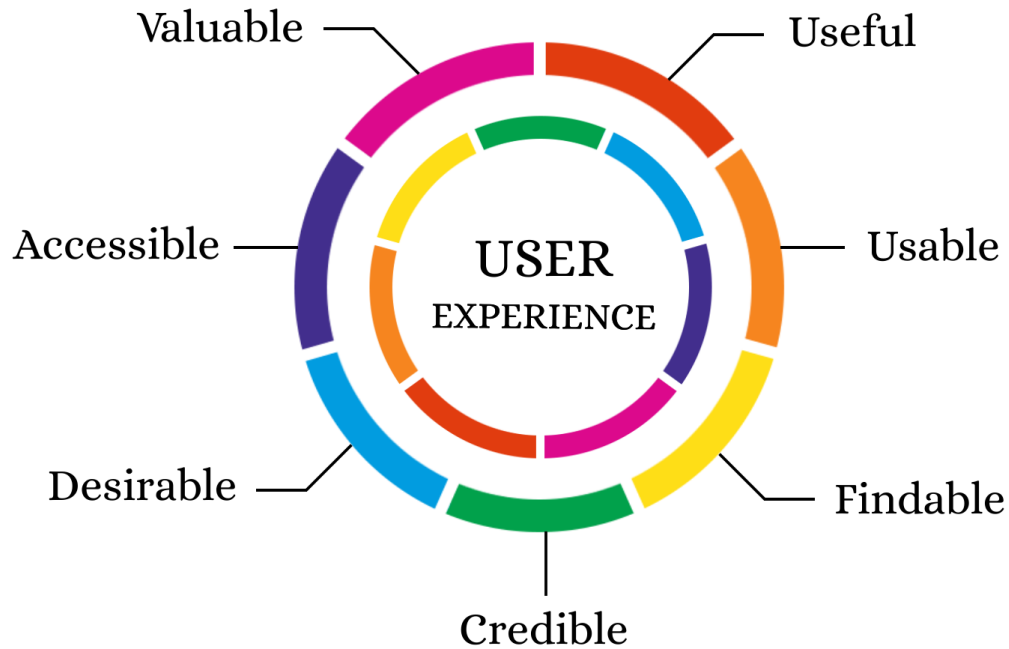
Organising, documenting data with an ELN

How to choose an ELN?

MUST HAVES

UX experience

USER FRIENDLY



Source: <https://uxdesign.cc/>



UHASSELT

KNOWLEDGE IN ACTION



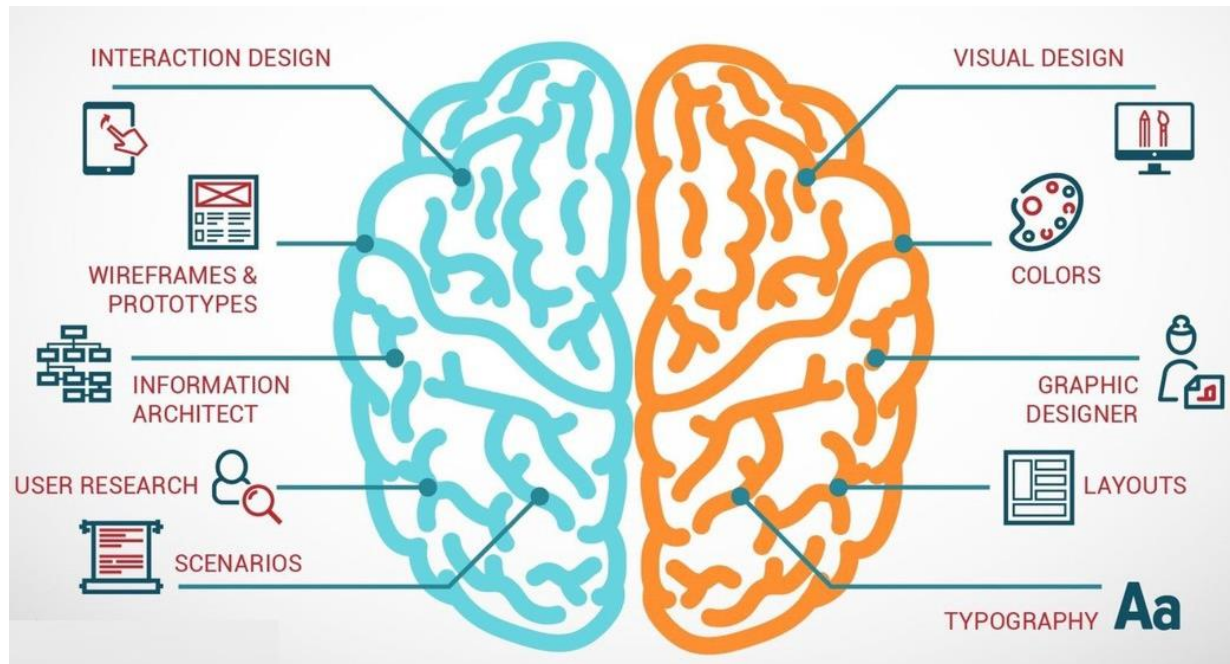
Organising, documenting data with an ELN

How to choose an ELN?

MUST HAVES

UX experience

UI interface





Organising, documenting data with an ELN

How to choose an ELN?

MUST HAVES

General functionalities

- ✓ **Link/ Export standard files**
- ✓ **Search/ Sort**
- ✓ **Equations**
- ✓ **Data backup**
- ✓ **Automatic Metadata creation**
- ✓ **Barcodes reading**





Organising, documenting data with an ELN

How to choose an ELN?

MUST HAVES

IT functionalities

- ✓ **Version control**
- ✓ **Secure storage location**
 - ✓ **Access**
 - ✓ **Data protection**
 - ✓ **Automatic backup**
 - ✓ **Storage of large data file**
- ✓ **Platform independent (ex: Webbased)**
- ✓ **Storage locations compatible local data storage regulations e.g. Europe**





Organising, documenting data with an ELN

How to choose an ELN?

MUST HAVES

Collaboration

- ✓ **Supervisory access/ accounts**
- ✓ **Possibility to share data**





Organising, documenting data with an ELN

How to choose an ELN?

SHOULD HAVES

General/ IT/ Collaboration/ Integration

- ✓ **Make annotation**
- ✓ **Zoom(resize) handwriting/drawings**
- ✓ **Create project templates
(No need to rewrite protocols!)**
- ✓ **Create project workflows**
- ✓ **Local & Cloud storage**
- ✓ **Single sign-on**
- ✓ **Possible to work offline**
- ✓ **Synchronisation between devices**
- ✓ **Audit trail (Validation)**
- ✓ **Time stamping (Valorisation)**
- ✓ **Creation of Persistent identifiers
(Findability)**
- ✓ **Linking protocols to experiments/data**
- ✓ **Connectivity with agenda, image tool**
- ✓ **Integration between active-archive data**
- ✓ **Integration of database**





Organising, documenting data with an ELN

How to choose an ELN?

NICE TO HAVE

- ✓ Write/audio to text function
- ✓ Bibliographic management
- ✓ Integration ORCID-id
- ✓ What to keep filter (prevent redundant data)
- ✓ Lab inventory

General

- ✓ Drawing tools
- ✓ Live experiment (Data capture real-time)
- ✓ Basic experimental tools:
 - ✓ Timers
 - ✓ Thermometers
 - ✓ Calculator
- ✓ Free
- ✓ Cloud-based

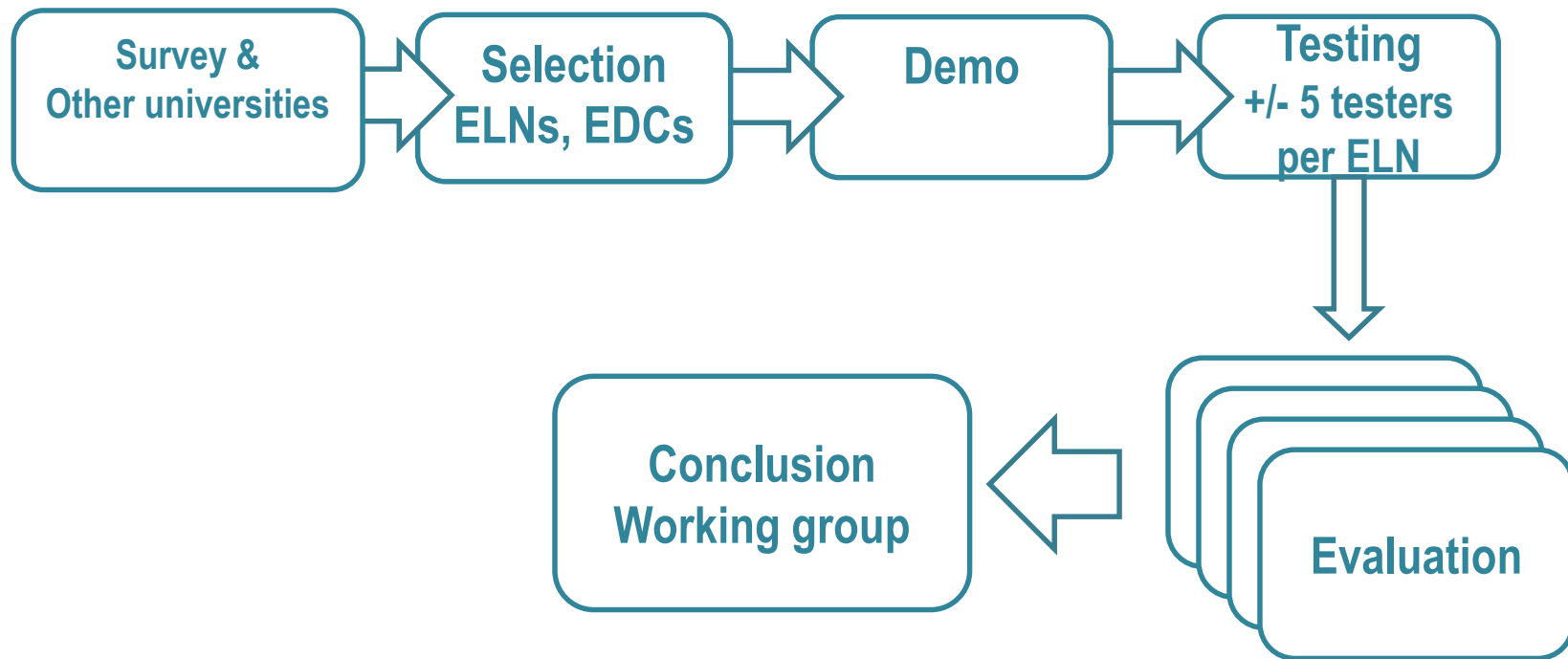




Organising, documenting data with an ELN

Methodology: Testing ELNs @UHasselt






+/- 1 month testing - 25 participants





Organising, documenting data with an ELN

Functionalities: Discipline specific/ Generic






ELN/ EDC	Discipline	Functionalities
 castor	→ Biomedical	→ Clinical trials
 SignalsNotebook Individual + Standard	→ Chemical	→ Chemical structures
 Benchling	→ Life sciences	→ DNA and protein creation
 OSF	→ General	→ General
 eLabFTW	→ General	→ General





Organising, documenting data with an ELN

Functionalities: Free version / Open source/ Data storage

ELN/ EDC	Free version	Hosting - Data storage
 castor	→ No	→ Included
 SignalsNotebook Individual + Standard	→ No	→ Included
 Benchling	→ Yes / No	→ Free version 10 GB Connection Google drive
 OSF	→ Yes / No (OS)	→ Free version 5 GB Connection Google drive
 eLabFTW	→ Yes (OS)	→ Not included

OSF = (Open science framework)
OS= Open source





Organising, documenting data with an ELN

Focus points testing

- ✓ Installation/ Registration
- ✓ UI (intuitively)
- ✓ Collaboration
- ✓ Access control
- ✓ Support training

*All scored very well on all of these points.
Some with more collaboration, access control possibilities*



Organising, documenting data with an ELN

Testing highlighted outcomes



Easy follow-up for Clinical trials



Individual + Standard



Chemical structures

Standard:

- ✓ Extended access control
- ✓ Task assignment
- ✓ Multidisciplinary libraries



DNA and protein creation/ simulation included



Many add-ons
Open source



eLabFTW



Many add-ons
Open source
Very intuitive
Leave comments for other researchers





How to choose an ELN?

GENERAL

Consider main goal

- ✓ Generic
- ✓ Discipline specific

Main technical features

Price

- ✓ Usage
- ✓ Hosting
- ✓ Data storage included

Open source





How to choose an ELN?

SPECIFIC

- ✓ **Your institution**

- ✓ Licences
- ✓ IT Support

- ✓ **Other user experiences, usage**

(In your institution, other institutions, european level)

On level of

- ✓ Installation/Activation
- ✓ User friendly (Intuitive)
- ✓ Collaboration and Access control
- ✓ Training, Support





Organising, documenting data with an ELN

How to choose an ELN?

Check

- ✓ Functionalities JISC
Consider what you find important





Takeaways

- ✓ Importance of organising and documenting data
- ✓ Prevent data loss
 - ✓ Make data **findable** by correct file naming, versioning, folder structure
 - ✓ Make data **understandable** by using readme files, other data documentation (folder structure, file naming convention, methodology, protocols, ontologies...) and automated metadata,...



Organising, documenting data with an ELN

Takeaways

- ✓ What is an ELN
- ✓ How to choose the correct ELN?

Consider

- ✓ Your main goal

Check

- ✓ Your institutions and other institutions
- ✓ Technical features
- ✓ Extra Functionalities (JISC)





Organising, documenting data with an ELN

Interesting links

ARTICLES

Organizing and documenting

Article: [Organizing files and folders](#) (Wageningen University)

Article: [Readme-files: a Template and an example](#) (Harvard Medical School)

Article: [A guideline for reporting experimental protocols in life sciences](#) (PeerJ.com) ^{cc}

ELNs

Article: [Turning the page on paper notebooks creates a digital dilemma](#) (JISC)

Article: [How to pick an electronic laboratory notebook](#) (Nature.com)

Article: [Keep calm and go paperless](#) (openworking 4TU researchData en TUDelft)

LISTS-MATRIXES

List: JISC-Glasgow [list of requirements](#) (must haves, should haves, could haves)

Matrix: [Harvard comparison grid](#), [Cambridge comparison grid](#)





Organising, documenting data with an ELN

Contact

hanne.vlietinck@uhasselt.be

Datasteward





Subjects

- ✓ **Organising and documenting data (basics)**
- ✓ **ELN (basics)**
(Hanne Vlietinck, Data Steward, U Hasselt)

- ✓ **Demo Benchling**
(Jonas Delva, PhD Student Life Sciences, UGent)

- ✓ **Demo ELabFTW**
(Ing. Nicolas Carpi, Engineer and Founder ELabFTW, Institute Curie Paris)

- ✓ **Experience from a PhD student (ELabFTW)**
(Ing. Robbe Breugelmans, PhD student U Hasselt)



Subjects

- ✓ **Organising and documenting data (basics)**
- ✓ **ELN (basics)**
(Hanne Vlietinck, Data Steward, U Hasselt)

- ✓ **Demo Benchling**
(Jonas Delva, PhD Student Life Sciences, UGent)

- ✓ **Demo ELabFTW**
(Ing. Nicolas Carpi, Engineer and Founder ELabFTW, Institute Curie Paris)

- ✓ **Experience from a PhD student (ELabFTW)**
(Ing. Robbe Breugelmans, PhD student U Hasselt)



Subjects

- ✓ **Organising and documenting data (basics)**
- ✓ **ELN (basics)**
(Hanne Vlietinck, Data Steward, U Hasselt)
- ✓ **Demo Benchling**
(Jonas Delva, PhD Student Life Sciences, UGent)
- ✓ **Demo ELabFTW**
(Ing. Nicolas Carpi, Engineer and Founder ELabFTW, Institute Curie Paris)
- ✓ **Experience from a PhD student (ELabFTW)**
(Ing. Robbe Breugelmans, PhD student U Hasselt)

