

Research data management and ELNs

Who am I



Nicolas CARPI

- Worked for 15 years in a cell biology research lab
- Creator of **eLabFTW** electronic lab notebook
- Founder and President of **Deltablot** company



Deltablot

What's an ELN?

ELN = Electronic Laboratory Notebook

- Vastly superior to paper notebooks
- First place where research data is referenced (close to the bench)
- Daily tool of any researcher in any branch of Science

What's FAIR



100% FAIR is utopia

- Research is messy by design
- Labs are highly dynamics, people come and go all the time
- Improving the FAIRness MUST NOT be a hassle and should be as automatic/transparent as possible

=> Using a progressive enhancement approach is the most likely to be successful

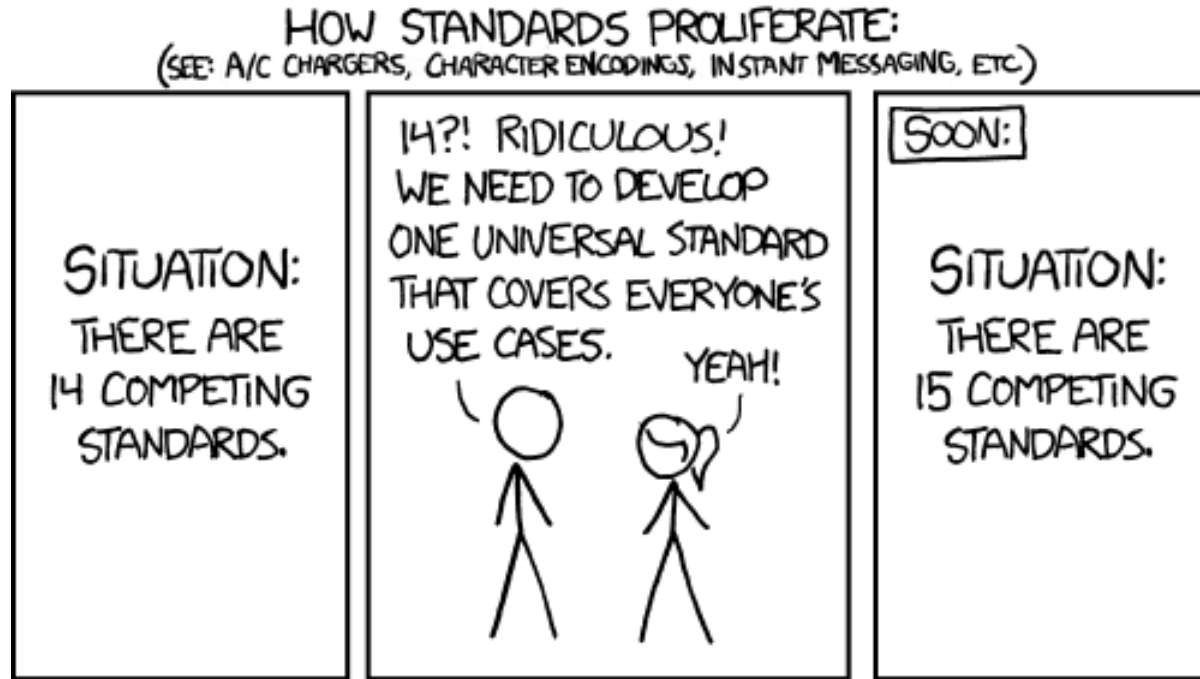
First step is to get the data in the ELN

- Make it easy to input data
- It's okay if the data doesn't have 184 ontology defined fields
- It's okay if it's incomplete, at least we have something
- Now we can search for it : that's the F

Making it Accessible

- That's the job of a data repository, and should happen after publication
- eLabFTW's approach to research data is private by default

Making it interoperable



Making it interoperable



<https://the.elnconsortium.org>

We need to describe research data in a way that any program can understand.

The ELN Consortium: we'll use RO-Crates and ZIP archives. => re-use existing technology!

RO-Crate contains *ro-crate-metadata.json* using **specific terminology** to describe the content of the archive.

The ELN file format is officially recognized by IANA (*application/vnd.eln+zip*).

Making it interoperable

How it works:

- eLabFTW: I use *date*
- Kadi4Mat: I use *start_date*
- PASTA: I use *experiment_date*
- SampleDB: I use *main_datetime*

Problem: we all have our vocabulary/format for different aspects of the data

Solution: common vocabulary that we can agree upon

Making it interoperable

Use common vocabulary

dateCreated

A Schema.org Property

Thing > Property :: dateCreated

The date on which the CreativeWork was created or the item was added to a DataFeed.

Values expected to be one of these types

- Date
- DateTime

Used on these types

- CreativeWork
- DataFeedItem

Sub-properties

- legislationDate

dateModified

A Schema.org Property

Thing > Property :: dateModified

The date on which the CreativeWork was most recently modified or when the item's entry was modified.

Values expected to be one of these types

- Date
- DateTime

Used on these types

- CreativeWork
- DataFeedItem

Making it interoperable

Now everyone uses DateTime in ISO8601 format.

We can import/export from different software.

Want to share a sample? Share also the .eln so the history of its creation is not lost!

Want to publish a project along with your scientific article? Export all the related entries in a .eln and push that to the data repository which can interpret all the metadata correctly.

ChemicalSubstance example

ChemicalSubstance

A Schema.org Type

This term is in the "new" area - implementation feedback and adoption from applications and websites can help improve our definitions.

Thing > [BioChemEntity](#) > [ChemicalSubstance](#)

[\[more...\]](#)

A chemical substance is 'a portion of matter of constant composition, composed of molecular entities of the same type or of different types' (source: [ChEBI:59999](#)).

Property	Expected Type	Description
Properties from ChemicalSubstance		
chemicalComposition	Text	The chemical composition describes the identity and relative ratio of the chemical elements that make up the substance.
chemicalRole	DefinedTerm	A role played by the BioChemEntity within a chemical context.
potentialUse	DefinedTerm	Intended use of the BioChemEntity by humans.
Properties from BioChemEntity		
associatedDisease	MedicalCondition or PropertyValue or URL	Disease associated to this BioChemEntity. Such disease can be a MedicalCondition or a URL. If you want to add an evidence supporting the association, please use PropertyValue.
bioChemInteraction	BioChemEntity	A BioChemEntity that is known to interact with this item.
bioChemSimilarity	BioChemEntity	A similar BioChemEntity, e.g., obtained by fingerprint similarity algorithms.
biologicalRole	DefinedTerm	A role played by the BioChemEntity within a biological context.
funding	Grant	A Grant that directly or indirectly provide funding or sponsorship for this item. See also ownershipFundingInfo . Inverse property: fundedItem
hasBioChemEntityPart	BioChemEntity	Indicates a BioChemEntity that (in some sense) has this BioChemEntity as a part. Inverse property: isPartOfBioChemEntity
hasMolecularFunction	DefinedTerm or PropertyValue or URL	Molecular function performed by this BioChemEntity; please use PropertyValue if you want to include any evidence.
hasRepresentation	PropertyValue or Text or URL	A common representation such as a protein sequence or chemical structure for this entity. For images use schema.org/image .

Making it interoperable



Result:

- export from one ELN to another
- export from an ELN to a data repository
- export from anything to anything that can open a zip and read JSON

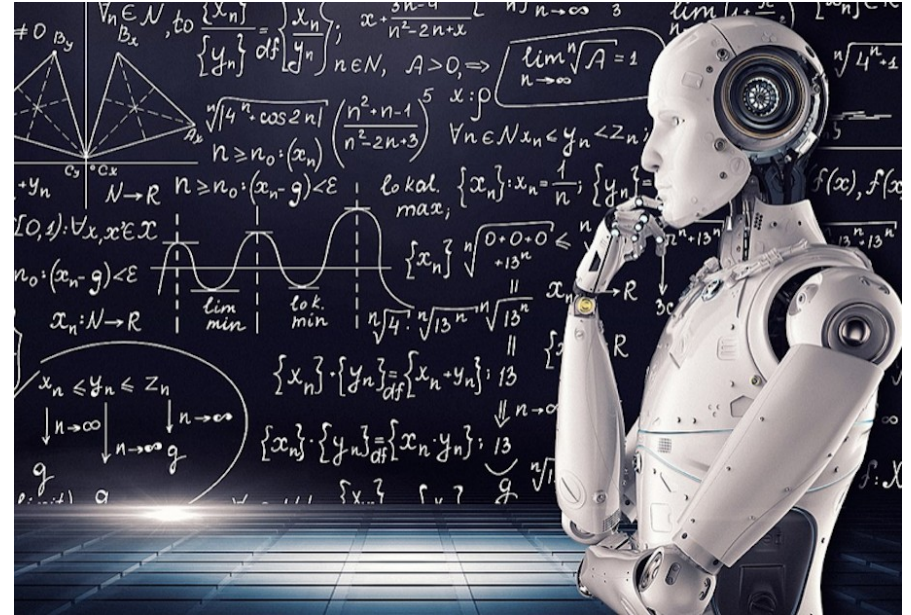
<https://the.elnconsortium.org>

Making it reusable

The research data is now conveniently annotated

More importantly: the experiment that produced that data is available with all the conditions and information needed to reproduce it

We can do some machine learning AI magic on it!



Next steps

Improve automatic metadata extraction from raw data

Improve ELN file format comprehension from Dataverse

Increase ELN file format support across the research software space

Use more precise ontologies

Links

- eLabFTW ELN: <https://www.elabftw.net>
- The ELN Consortium: <https://the.elnconsortium.org>
- This presentation:
<https://github.com/elabftw/elabpdf/blob/master/talks/2023-08-14-ACS/acs.pdf>

Scan for link to this presentation:

