

# From Searching to Publishing Scientific Information ENV-444

Vincenzo Palatella

EPFL  
27.10.2025



# Learning objectives

At the end of this training, you will be able to:

- Find academic literature
- Use a reference manager as Zotero
- Work collaborative on shared documents with colleagues
- Publish a dataset on Zenodo

## Part 1

- Introduction (5')
- Information retrieval (20')
  - Search strategy
  - From Google Scholar to OpenAlex
- Zotero (20')

## Part 2

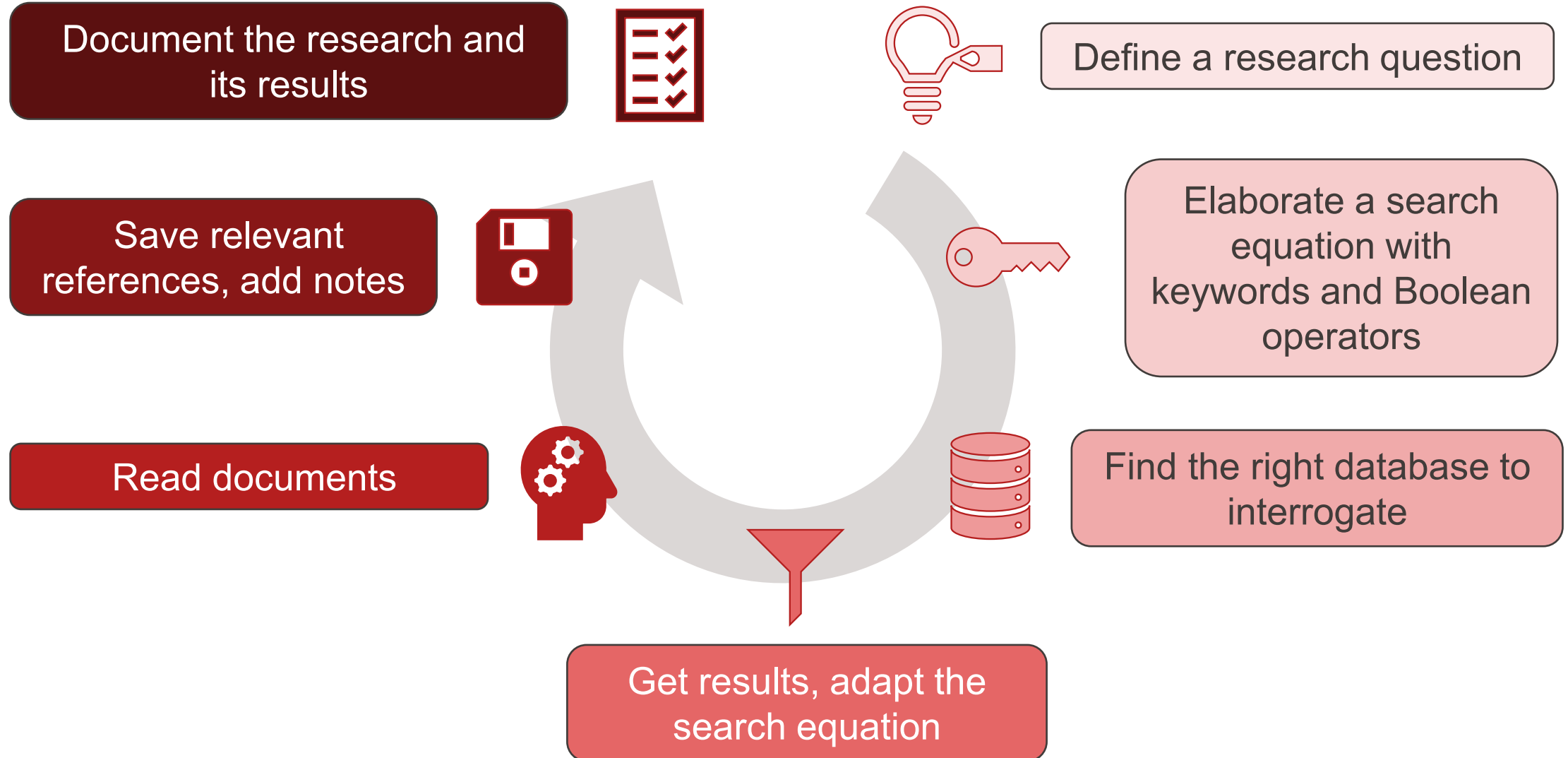
- Collaborative writing (35')
  - From Word to Overleaf
  - Let's get to work!
- Publishing a dataset (10')
  - Zenodo
  - DOI

## Break



# Information retrieval

# Information retrieval strategy



# Define research question(s)



- What's your research question?

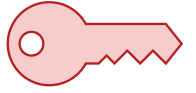
*What's the role of sugar in obesity?*

# The research grid



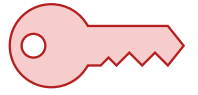
Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>			
Search for <b>synonyms</b> and <b>related terms</b>			
Create the <b>search equation</b>			
Search 2 databases and compare their <b>results</b>			
<b>Refine</b> your search equation			

# The research grid



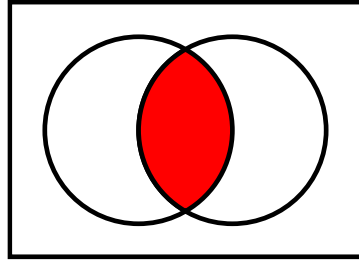
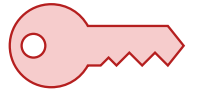
Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>	Sugar	Obesity	Role
Search for <b>synonyms</b> and <b>related terms</b>			
Create the <b>search equation</b>			
Search 2 databases and compare their <b>results</b>			
<b>Refine</b> your search equation			

# The research grid

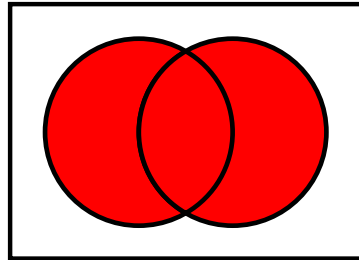


Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>	Sugar	Obesity	Role
Search for <b>synonyms</b> and <b>related terms</b>	Glucose Fructose	Metabolism Eating disorder	
Create the <b>search equation</b>			
Search 2 databases and compare their <b>results</b>			
<b>Refine</b> your search equation			

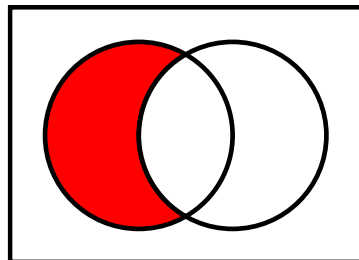
# Boolean operators and more



**AND** to combine key concepts



**OR** to add alternative key concepts



**NOT** to exclude concepts

**( )**

Group the keywords

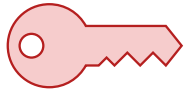
**\*** or **?**

Jolly character

**<< >>** or **" "**

Exact expression

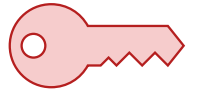
# The research grid



Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>	Sugar	Obesity	Role
Search for <b>synonyms</b> and <b>related terms</b>	Glucose Fructose	Metabolism Eating disorder	<b>AND</b>
Create the <b>search equation</b>			
Search 2 databases and compare their <b>results</b>			
Refine your search equation			

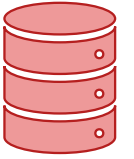


# The research grid



Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>	Sugar	Obesity	Role
Search for <b>synonyms</b> and <b>related terms</b>	Glucose Fructose	Metabolism Eating disorder	
Create the <b>search equation</b>	(sugar* OR glucose OR fructose) AND (obesity OR metabolis* OR "eating disorder") AND role		
Search 2 databases and compare their <b>results</b>			
<b>Refine</b> your search equation			

# Find the right database to interrogate



## Web of Science™



Scopus

Google Scholar

OpenAlex

# Web of Science™

[www.webofscience.com](http://www.webofscience.com)

- **Property of:** Clarivate Analytics
- **Access:** subscription-based (paid by the library)
- **Curation:** human (curated selection of high-impact journals and conference series)
- **Coverage:** ~79 M records, mostly peer-reviewed



# Scopus

[www.scopus.com](http://www.scopus.com)

- **Property of:** Elsevier
- **Access:** subscription-based (paid by the library)
- **Curation:** human (curated selection of high-impact journals and conference series)
- **Coverage:** ~78 M records, mostly peer-reviewed

# Google Scholar

[scholar.google.com](https://scholar.google.com)

- **Property of:** Google
- **Access:** free (with or without a Google account)
- **Curation:** automated indexing
- **Coverage:** ~ 160 million records ( $\pm 10\%$ ), not always peer-reviewed (grey literature)



[openalex.org](https://openalex.org)

- **Property of:** Our Research
- **Access:** free and completely open source
- **Curation:** automated indexing
- **Coverage:** ~ 240 million works

# The research grid



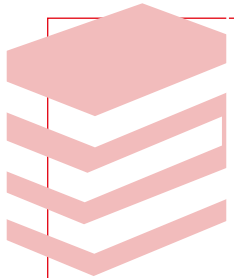
Define a <b>subject</b> and transform it into a <b>question</b>	What's the role of sugar in obesity?		
Extract the main <b>concepts</b> and get <b>keywords</b>	Sugar	Obesity	Role
Search for <b>synonyms</b> and <b>related terms</b>	Glucose Fructose	Metabolism Eating disorder	
Create the <b>search equation</b>	(sugar* OR glucose OR fructose) AND (obesity OR metabolis* OR "eating disorder") AND role		
Search 2 databases and compare their <b>results</b>	<b>OpenAlex</b> 21'660 results	<b>Scopus</b> 115'000 results	
<b>Refine</b> your search equation	Too many results? Use <b>filters</b> / <b>add concept(s)</b> Too few results? <b>Remove concept(s)</b> / <b>add synonym(s)</b>		

# What about LLMs ?

Message ChatGPT...



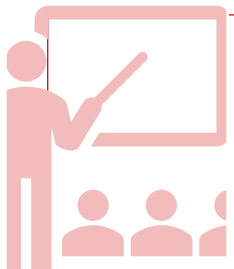
ChatGPT can make mistakes. Consider checking important information.



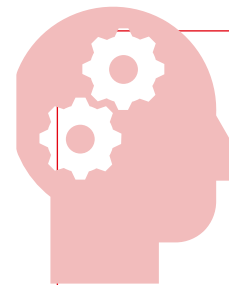
Never use it as only source of information



Be transparent about its use in your works



Always follow the instructions provided by your professor



When possible, use tools trained with academic corpus - [Scite](#), [Elicit](#), [Consensus](#), with trackable references



# Introduction to Zotero

# Why a reference manager?

You can expect a reference manager to help you:

- Adding/importing references very easily
- Creating a bibliography very easily
- Citing while you write a paper

All these technical tasks should be done by the software but remember: **you are the brain.**

This means that you should:

- Check if the reference data are correct and complete
- Read papers
- Sort references
- Select relevant references to add in the bibliography

# Why Zotero?



User-friendly and powerful

Free and Open Source



Modular

Active developers community





# How does a Zotero library look like?

From Searching to Publishing Scientific Information

The screenshot displays the Zotero desktop application. The main window shows a list of items in a table with columns: Title, Cre..., Item Type, Year, Publisher, Publication, Date Added, and Date Modified. The selected item is 'Contributions to the Theory of Riemann Surfaces. (AM-30), Volume 30' by Ahlfors, Lars Valerian, published in 1953 by Princeton University Press.

The right-hand pane shows the 'Info' section for the selected item, providing detailed metadata:

- Item Type: Book
- Author: Ahlfors, Lars Valerian
- Author: Calabi, E.
- Author: Morse, Marston
- Author: Sario, Leo
- Author: Spencer, Donald Clayton
- Series: Annals of Mathematics Studies
- Series Number: 28
- Volume:
- # of Volumes:
- Edition: Princeton University Press on dema ...
- Place: Princeton, NJ
- Publisher: Princeton University Press,
- Date: 1953
- # of Pages: 276
- Language: eng
- ISBN: 978-1-4008-2837-1
- Short Title:
- URL:
- Accessed:
- Archive:
- Loc. in Archive:
- Library Catalog: slsp-epfl.primo.exlibrisgroup.com
- Call Number:
- Rights:
- Extra: Book Title: Contributions to the Theory of Riemann Surfaces. (AM-30), Volume 30; DOI: 10.1515/9781400828371
- Date Added: 11/18/2021, 11:04:17 AM
- Modified: 11/18/2021, 11:04:17 AM

A large 'Demo' watermark is overlaid on the right side of the interface.

# Using Zotero with your word processor

- Bibliographies
- In-text citations and references

- What a citation style is?  
It's a set of rules on how to cite sources in academic writing.
  - Which metadata are displayed
  - How metadata are displayed

## American Chemical Society

(1) Zhou, J.; Xu, G.; Mu, C. Analysis of a Pseudo-Parabolic Equation by Potential Wells. *Annali di matematica pura ed applicata* **2021**, *200* (6), 2741–2766. <https://doi.org/10.1007/s10231-021-01099-1>.

## Chicago 17<sup>th</sup> ed. (full note)

Zhou, Jun, Guangyu Xu, and Chunlai Mu. 'Analysis of a Pseudo-Parabolic Equation by Potential Wells'. *Annali Di Matematica Pura Ed Applicata* 200, no. 6 (2021): 2741–66. <https://doi.org/10.1007/s10231-021-01099-1>.

# Creating an account

- Creating an account is not compulsory but:
  - It allows to have your library synchronized on the cloud (back-up)
  - It allows to create and work in groups
  
- Registering with an @epfl.ch email address will give you access to unlimited storage of full-texts (instead of 300MB)





# Collaborative writing

- Generic tools: MS Word (with O365), Google Docs
- Specialized in scholarly/scientific writing: Authorea, Overleaf

# Working together

Feature	Microsoft Word (online/desktop with OneDrive)	Google Docs	Authorea	Overleaf
<b>Primary purpose</b>	General-purpose word processing; strong formatting and layout controls	Cloud-native real-time editing of rich-text documents	Collaborative authoring of scholarly manuscripts	Real-time LaTeX editor aimed at scientific papers, theses, and technical reports
<b>Supported file formats</b>	Editing: doc/.docx, rtf Import: tex (LaTeX) - limited Export: pdf	Editing: gdoc, docx, odt Import/Export: pdf, rtf, txt	Editing: tex (LaTeX), md (Markdown) Import: docx Export: pdf	Editing: tex (LaTeX) Import: GitHub project Export: pdf, html
<b>Zotero integration</b>	Yes	Yes	Yes (also with native reference manager)	Yes
<b>Templates for scholarly publishing</b>	Limited	Very limited	Nature, IEEE, PLOS, etc.	Extensive collection (Springer, ACM, IEEE, arXiv, etc.)
<b>Pricing model</b>	Free for EPFL students	Free with a Google account	Free tier and paid plans (start at \$12 /mo)	Free tier, free Premium account when registering with an @epfl.ch email address



[word.cloud.microsoft](https://word.cloud.microsoft)

- **Property of:** Microsoft
- **Access:** paid via O365 (free for EPFL students)
- + widely adopted
- + collaborative
- not designed for academic writing (limited availability of templates, flexibility)



# Google Docs

[scholar.google.com](https://scholar.google.com)

- **Property of:** Google
- **Access:** free (with a Google account)
- + widely adopted
- + collaborative
- not designed for academic writing (very limited availability of templates, flexibility)

# AUTHOREA

[www.authorea.com](http://www.authorea.com)

- **Property of:** Authorea
- **Access:** free tier and premium features
- + flexible (LaTeX + Markdown)
- + vast collection of templates
- + collaborative
- less adopted



[www.overleaf.com](http://www.overleaf.com)

- **Property of:** Google
- **Access:** free tier and premium features (free for EPFL students signing in with @epfl.ch email address)
- + widely adopted in the academic world
- + basic knowledge of LaTeX needed
- + collaborative
- not designed for academic writing

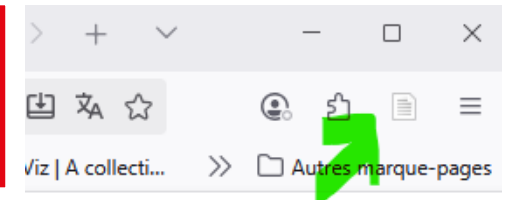
# Let's get to work now!

10'

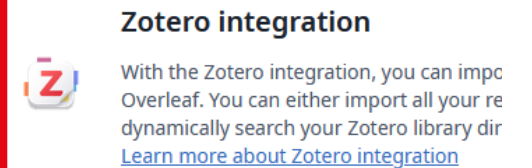
1 On a subject of your choice, search for academic sources in one – or more – of the tools we saw (Google Scholar, Scopus, Web of Science, OpenAlex).



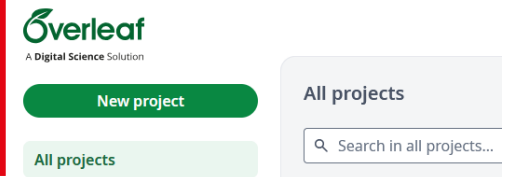
2 Create a Zotero library and use the browser connector to collect 3 references (articles, book chapter, conference papers...); add them to the library you've just created.



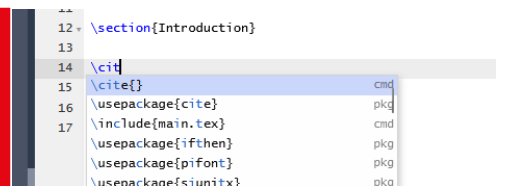
3 In Overleaf's settings, link your account with the Zotero one.



4 Create a new Overleaf project and share it with the colleague next to you;



5 Add your Zotero library as references and create a citation in the text.





# Publishing a dataset

The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo and a menu icon. Below the header, the EPFL logo and name "EPFL - École Polytechnique Fédérale de Lausanne" are displayed. A green "New upload" button is visible. The navigation bar includes "Search", "Requests", "Members", "Settings", "Curation policy", and "About". The search results section shows "352 results found" and a "Sort by" dropdown menu set to "Newest". Two dataset entries are visible:

- January 4, 2023 (v2.2)** Dataset Open  
**Arcangelo Corelli - Sonate a tre (A corpus of annotated scores)**  
 Johannes Hentschel ; Fabian C. Moss ; Markus Neuwirth   
 Arcangelo Corelli - Trio Sonatas (A corpus of annotated scores) This corpus of annotated MuseScore files has been created within the DCML corpus initiative and employs the DCML harmony annotation standard. It was released together with and as part of the "workflow paper" Hentschel, J., Moss, F. C., Neuwirth, M., & Rohrmeier, M. A. (2021). A...  
 Uploaded on October 25, 2023  
 4 more versions exist for this record 53 0
- October 2, 2023 (v4)** Dataset Open  
**Rethinking pose estimation in crowds: overcoming the detection information bottleneck and ambiguity**  
 Zhou, Mu ; Stoffl, Lucas ; Mathis, Mackenzie W   
 ##### Rethinking pose estimation in crowds: overcoming the detection information-bottleneck and ambiguity, ICCV 2023 ##### Authors: Zhou, Mu and Stoffl, Lucas and Mathis, Mackenzie Weygandt and Mathis, Alexander Affiliation: EPFL Date: October, 2023 Here we provide neural networks weights for the best models in our article "R...  
 Uploaded on October 25, 2023  
 3 more versions exist for this record 102 26

[zenodo.org](https://zenodo.org)

- Free
- Developed by CERN, paid by EU, hosted in European servers
- Max 50GB/dataset
- Unlimited number of datasets
- Automated DOI assignment
- Publishing metadata
- A lively EPFL community:

[zenodo.org/communities/epfl/](https://zenodo.org/communities/epfl/)

Demo

# How to publish a dataset to Zenodo

1. Go to <https://sandbox.zenodo.org/>
2. Sign up to create a Zenodo account
3. Choose New upload in the menu (top right of the page)
4. You can select the “EPFL” community where to submit your record ([curation policy](#))
5. Get a DOI
6. Choose “dataset” as resource type, word a title, specify a publication date
7. Add the creator (you are the only author)
8. Create an original text to describe your dataset
9. Choose an appropriate type of open license (Creative Commons Attribution 4.0 International).  
Check the information on these types of open license [here](#)
10. Add keywords
11. Add a field (domain of research)
12. Complete the other elements required and publish your dataset

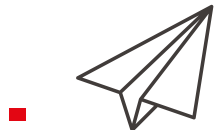
# Question time





Your feedback is important!

**Thank you!**



[formations.bib@epfl.ch](mailto:formations.bib@epfl.ch)



[library.epfl.ch](http://library.epfl.ch)  
[Our trainings](#)



[Book a librarian](#)