

Nerd approaches to scientific literature

Or how to navigate complexity without losing yourself

Giovanni Spitale, MA, PhD Student @IBME, UZH
giovanni.spitale@ibme.uzh.ch



**University of
Zurich**^{UZH}

Institute of Biomedical Ethics
and History of Medicine

AIMS



**University of
Zurich** ^{UZH}

Institute of Biomedical Ethics
and History of Medicine

1. Discuss the issue of publications' proliferation and present new techniques to cope with it;
2. Present what is open science and how is it reshaping our work;
3. Present some clever strategies to find an appropriate journal;
4. Maybe have some fun in the meanwhile.

GENERAL NOTES



**University of
Zurich** ^{UZH}

Institute of Biomedical Ethics
and History of Medicine

1. Hands on imperative! The best way to learn it is to try it out;
2. Expand your skillset – you’ll never know where life will bring you;
3. Challenge the status quo – even if often you’ll be miserably wrong;
4. There might always be a more clever or fastest way to do it;
5. Share it if it works, and spare other the hassle to reinvent the wheel

1. Tooo much

1. Too much

| Query | Meaning | Hits |
|--|--|--------|
| ("2020/10/01"[PDAT] : "2020/10/10"[PDAT]) | Every paper indexed in the last 10 days | 85,140 |
| ("2020/10/01"[PDAT] : "2020/10/10"[PDAT])AND covid[Title/Abstract] | Every paper indexed in the last 10 days mentioning «covid» in title or abstract | 5,823 |
| ("2020/10/01"[PDAT] : "2020/10/10"[PDAT])AND Coronaviridae[MeSH] | Every paper indexed in the last 10 days mentioning «Coronaviridae» in MeSH terms | 2,983 |

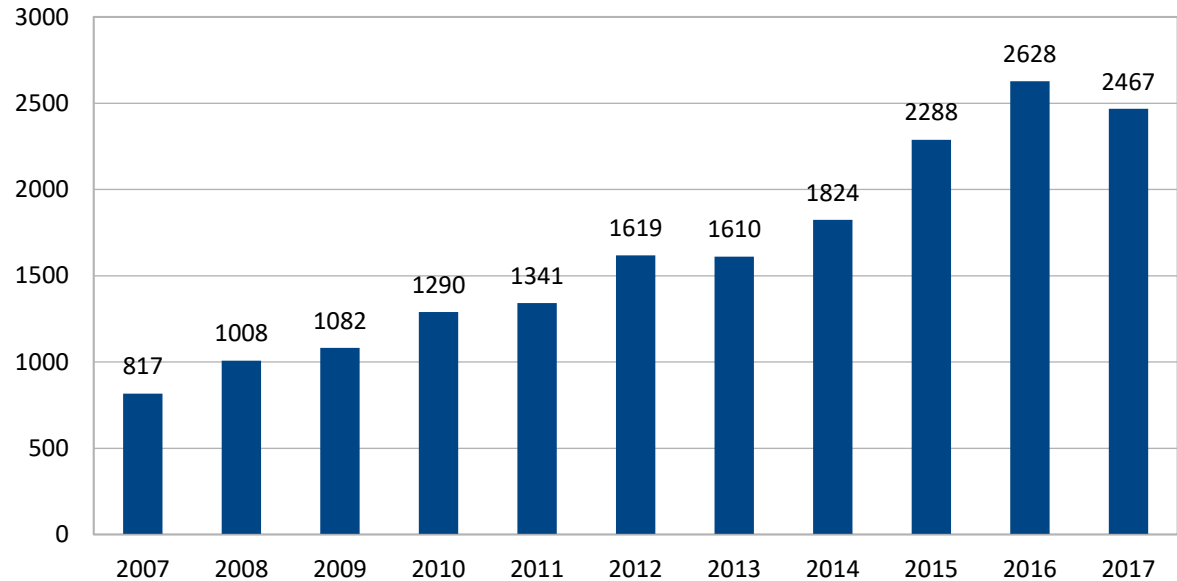
1. Too much

If instead of looking to the last ten days we consider ten years, the scenario becomes overwhelming: this graph displays the number of publications found on Web of Science with the query “end of life”:

TS=("end of life") AND PY=(2007-2017)

17974 results.

End of Life, publications 2007 - 2017



Ambiguity is baaad

While in medical publications exists a (quite) unambiguous, clear and largely accepted system to index medical subject headings (MeSH), there is no such a thing for ethics. It follows that often you don't find what you are looking for simply because it uses different keywords/synonyms.

As an example...

1. Too much

From Davies et al., 2015:

“Despite the increased prevalence of bioethics research that seeks to use empirical data to answer normative research questions, there is no consensus as to what an appropriate methodology for this would be”.

Davies, Rachel; Ives, Jonathan; Dunn, Michael (2015): A systematic review of empirical bioethics methodologies. In: BMC medical ethics, 16, p. 15. DOI: 10.1186/s12910-015-0010-3.

| Dialogical process | Combination of dialogical/consultative | Consultative process | Neither clearly dialogical nor consultative |
|--|--|--|---|
| <ul style="list-style-type: none">•Inter-ethics•Response evaluation hermeneutics•Moral experience hermeneutics•Moral conversation | <ul style="list-style-type: none">•Pragmatic hermeneutics•Deliberative democracy•Integrated empirical ethics | <ul style="list-style-type: none">•Encounters with experience•Phenomenological hermeneutics and wide reflective equilibrium•Wide reflective equilibrium and overlapping consensus•Network model with third person moral experience•Normative empirical reflective equilibrium•... | <ul style="list-style-type: none">•Interdisciplinary epoche•Ethics of public understanding•Micro-ethics•Oppositional collaboration•Complementary thesis•Distinct methodological collaboration•Phenomenological hermeneutics |
| 4 | 3 | 22 | 7 |

What to read first?

If we want to have a comprehensive understanding of a field of medical ethics, even a quite narrow one, we cannot read everything and retain the relevant information. Moreover, as proposed by Theodore Sturgeon (and recently endorsed by Daniel Dennett) “ninety percent of everything is crap”; or, in other words, we cannot be sure a priori that every piece of literature we retrieve is worth being read.

Therefore, while approaching a new field, we have two distinct problems:

1. (Assuming for the sake of discussion Sturgeon’s law as true) how can we reduce the amount of non relevant/interesting literature in our corpus, without wasting too much time and losing relevant information?
2. 10% of 5823 (covid papers) is still a lot of stuff to read. How can we approach such an amount of literature so that we have both a granular understanding of the single paper and an overall view of the main topics in the field?

1. Too much

Regarding problem one:

“The newer, the better” is an approach which is not viable in medical ethics.

Plenty of relevant literature is old, especially if we want to understand the development of a certain practice or issue over time;

“The most cited, the better” is a flawed approach in principle: it starts a positive feedback mechanism that leads to marginalize articles that might be relevant, but for some reason didn’t receive an initial burst of quotations at their publication (“reputation echo chamber”);

Other approaches (like “follow a specific tradition/approach”) are flawed in principle: we lose a global perspective on the field (“heritage echo chamber”).

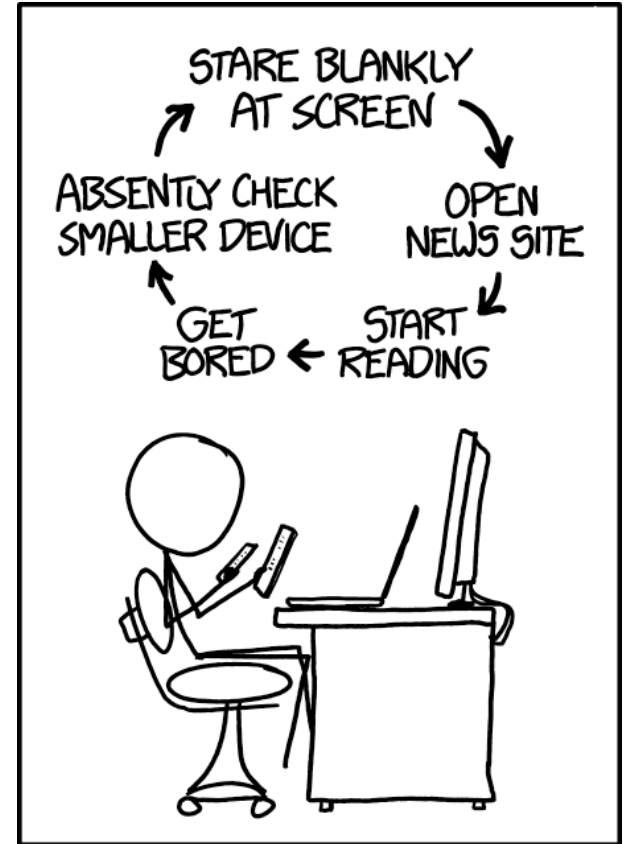
Solution: algorithmic approaches to literature retrieving → smart (and iterative) search strategies.

Regarding problem two:

“First in, first out, and read everything”: you end up with a massive amount of notes, precise on the single paper but lacking an overall picture;

“Read the abstract first, read the paper only if the abstract seems relevant”: you risk to arbitrarily miss relevant studies just because the abstract was not fancy enough.

Solution: algorithmic approaches to data → smart data manipulation.

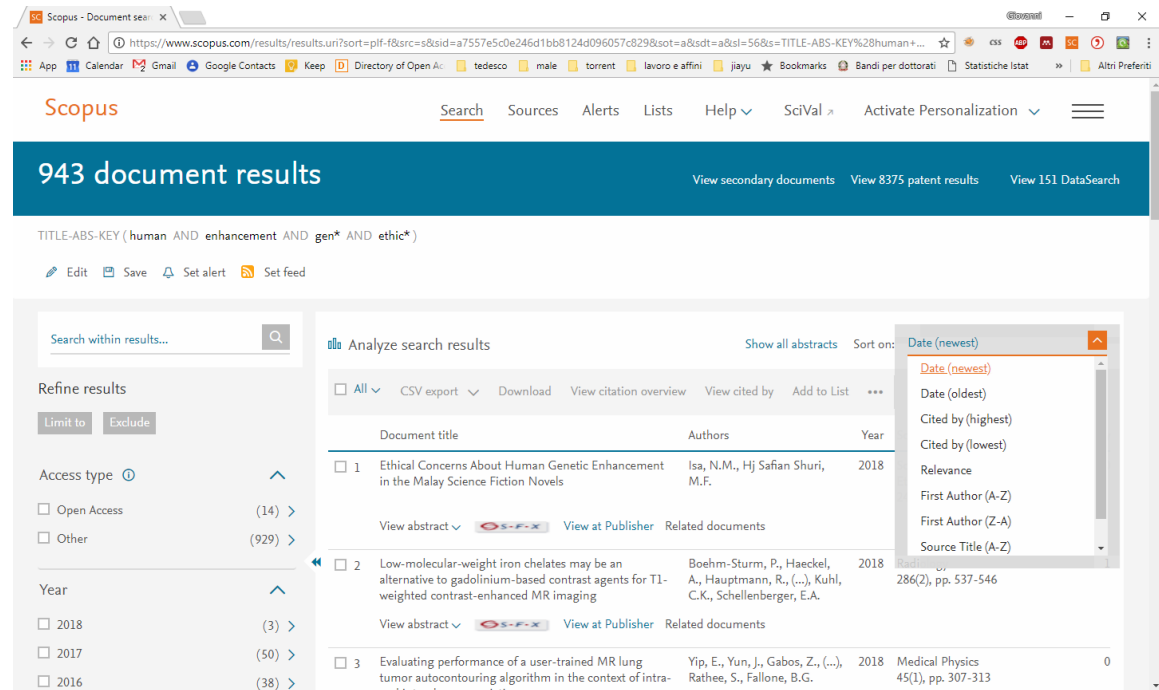


2. One ring to find them all

2. One ring to find them all

Query:
TITLE-ABS-KEY(human AND enhancement AND gen* AND ethic*)

Results:
Ordered by date, first the oldest ones



Scopus - Document search

943 document results

TITLE-ABS-KEY(human AND enhancement AND gen* AND ethic*)

Search within results...

Refine results

Limit to Exclude

Access type

Open Access (14)

Other (929)

Year

2018 (3)

2017 (50)

2016 (38)

Analyze search results

Show all abstracts Sort on: Date (newest)

Date (newest)

Date (oldest)

Cited by (highest)

Cited by (lowest)

Relevance

First Author (A-Z)

First Author (Z-A)

Source Title (A-Z)

| Document title | Authors | Year |
|---|--|------|
| 1 Ethical Concerns About Human Genetic Enhancement in the Malay Science Fiction Novels | Isa, N.M., Hj Safian Shuri, M.F. | 2018 |
| 2 Low-molecular-weight iron chelates may be an alternative to gadolinium-based contrast agents for T1-weighted contrast-enhanced MR imaging | Boehm-Sturm, P., Haeckel, A., Hauptmann, R., (...), Kuhl, C.K., Schellenberger, E.A. | 2018 |
| 3 Evaluating performance of a user-trained MR lung tumor autocontouring algorithm in the context of intra- | Yip, E., Yun, J., Gabos, Z., (...), Rathee, S., Fallone, B.G. | 2018 |

2. One ring to find them all

Export format:

CSV (comma separated values, basically a text file that can also be opened as an excel file).

Export fields:

Author, title, year, abstract and keywords (all the contentwise relevant information).

[Example](#)

Scopus - Document search

https://www.scopus.com/results/results.uri?sort=plf-f&src=s&sid=a7557e5c0e2461bb8124d096057c829&stot=a&stl=56&st=TITLE-ABS-KEY%28human+...

Scopus

Export document settings

You have chosen to export 943 documents

Select your method of export

MENDELEY RefWorks RIS Format CSV BibTeX Plain Text

EndNote, Reference Manager Excel ASCII in HTML

What information do you want to export?

| | | | | |
|--|---|---|---|--|
| <input type="checkbox"/> Citation information | <input type="checkbox"/> Bibliographical information | <input checked="" type="checkbox"/> Abstract & keywords | <input type="checkbox"/> Funding details | <input type="checkbox"/> Other information |
| <input checked="" type="checkbox"/> Author(s) | <input type="checkbox"/> Affiliations | <input checked="" type="checkbox"/> Abstract | <input type="checkbox"/> Number | <input type="checkbox"/> Tradenames & manufacturers |
| <input checked="" type="checkbox"/> Document title | <input type="checkbox"/> Serial identifiers (e.g. ISSN) | <input checked="" type="checkbox"/> Author keywords | <input type="checkbox"/> Acronym | <input type="checkbox"/> Accession numbers & chemicals |
| <input checked="" type="checkbox"/> Year | <input type="checkbox"/> PubMed ID | <input checked="" type="checkbox"/> Index keywords | <input type="checkbox"/> Sponsor | <input type="checkbox"/> Conference information |
| <input type="checkbox"/> EID | <input type="checkbox"/> Publisher | <input type="checkbox"/> Funding text | <input type="checkbox"/> Include references | |
| <input type="checkbox"/> Source title | <input type="checkbox"/> Editor(s) | | | |
| <input type="checkbox"/> volume, issue, pages | <input type="checkbox"/> Language of original document | | | |
| <input type="checkbox"/> Citation count | <input type="checkbox"/> Correspondence address | | | |
| <input type="checkbox"/> Source & document type | <input type="checkbox"/> Abbreviated source title | | | |
| <input type="checkbox"/> DOI | | | | |
| <input type="checkbox"/> Access Type | | | | |

Cancel Export

3 Evaluating performance of a user-trained MR lung tumor autocontouring algorithm in the context of intra- Yip, E., Yun, J., Gabos, Z., (...), 2018 Medical Physics 45(1), pp. 307-313 Rathee, S., Fallone, B.G.

2. One ring to find them all

Myers D.G., Schreiber F.B., Viel D.J.,
Effects of discussion on opinions concerning illegal behavior, 1974

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-0015977332&doi=10.1080%2f00224545.1974.9923074&partnerID=40&md5=b1bfbf199127b86b9223d6613134833c>

In an attempt to generalize recent research on the effects of group discussion and to further elucidate mechanisms responsible for discussion-induced response change, 15 groups of institutionalized males convicted of felony and 14 groups of college males responded, before and after discussion, to three choice dilemma items and three ethical-legal dilemmas. It was predicted that (a) the enhancement of mean initial tendency observed on choice dilemma items would extend to the ethical-legal dilemmas, and (b) shift to increased risk on both types of items would occur if, and only if, subjects tended to perceive themselves as initially riskier than their average peer. On the choice dilemmas (with a Likert type response format) only small nonsignificant shifts occurred. On the ethical-legal dilemmas both the inmate and the college samples significantly increased their preference for the legally deviant action following discussion, although initially perceiving themselves to favor it less than their peers. © 1974 Taylor & Francis Group, LLC.

alginate; behavior; ethics; injury; institutionalization; major clinical study; Attitude; Criminal Psychology; Ethics; Gambling; Group Processes; Human; Male; Self Concept",2-s2.0-0015977332

2. One ring to find them all

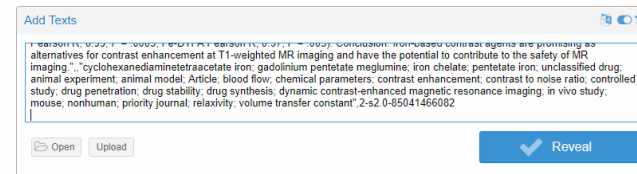
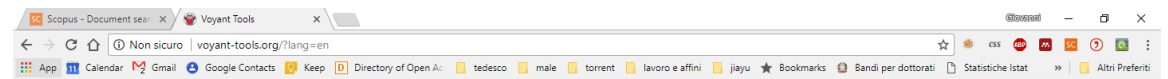
Voyant Tools:

“A web-based text reading and analysis environment. It’s designed to make it easy for you to work with your own text or collection of texts in a variety of formats, including plain text, HTML, XML, PDF, RTF, and MS Word”.

(Sinclair, Stéfan, Geoffrey Rockwell and the Voyant Tools Team. 2012. Voyant Tools (web application). <http://docs.voyant-tools.org/>).

- Free and open source;
- Well documented and easy to use;
- Runs both online or locally.

Example



Voyant Tools is a web-based reading and analysis environment for digital texts.



What's in the data?

Below the surface of “genetic human enhancement ethics” we have a plethora of terms giving hints on what is going on in the field:

Looking to the most frequent terms, we know that the question is considered “medical”, strictly connected with social issues, one of the main subfields is reproduction, and in the last years there has been a raise of terms like “care” and “public”. The trend of “moral” is increasing, while the trend of “ethics” is decreasing.

The trends of “gen*” (means: genetics, genetic, gene, ...) are decreasing, while the trends of “brain” and “neuro*”(means: neurology, neuroethics, neuroethic, neuroscience, ...) are increasing.

From search to research questions

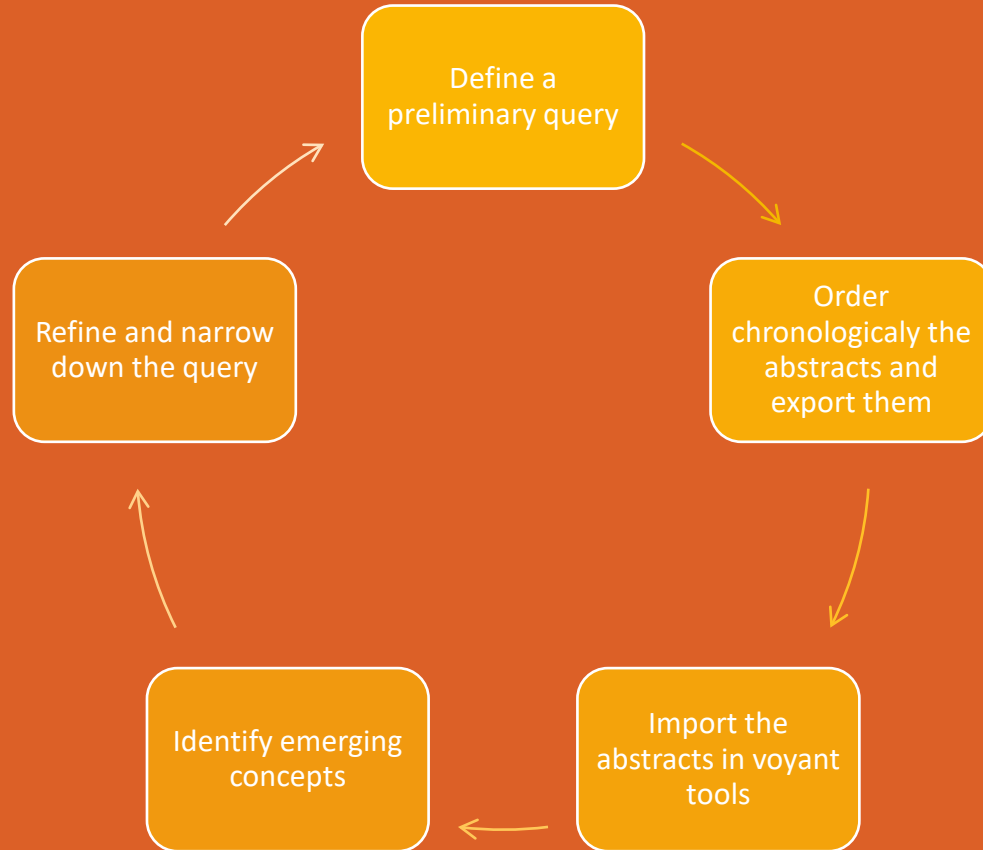
If we want to be traditional, we can start working on “Genetic human enhancement, reproduction and public policies: a care ethics perspective on the emerging social issues” → TITLE-ABS-KEY (human AND enhancement AND ethic* AND polic* AND soci* AND (reprod* OR child*)), 72 results;

If we want to surf the new wave, we can help ourselves with “Neuroenhancement: human brain improvement and moral challenges”. → TITLE-ABS-KEY(human AND enhancement AND (ethic* OR moral*) AND (brain or neuro*)), 434 results.

The process can be iterated in order to understand if our query actually captures what we are looking for and refine it...

... but still we have to retrieve and read all these papers.

Smart Iterative Search Strategies

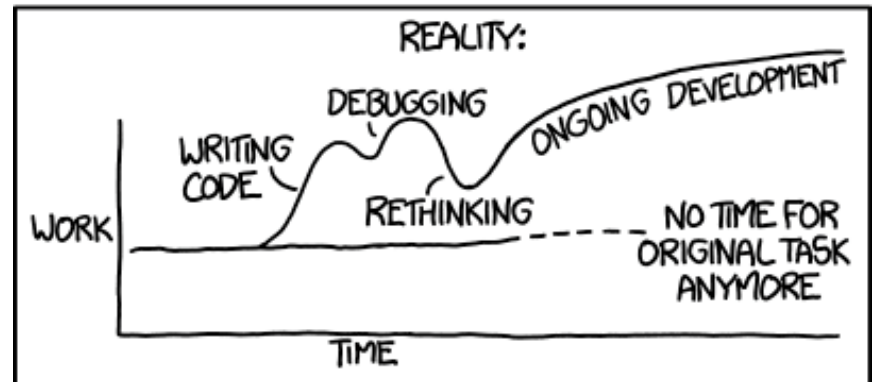
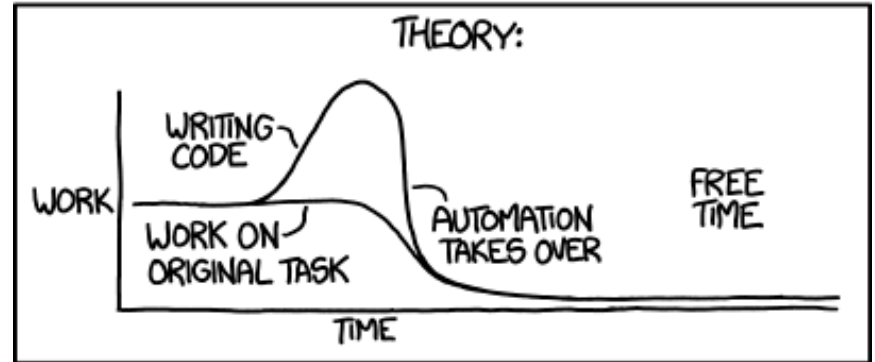


2. One ring to find them all

Exercise:

- Define a query that is relevant for your work
- Run it on a database of your choice
- Export the abstracts (ordered by year of publication)
- Explore their content in Voyant Tools
- Define a refined query

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



3. One ring to bring them all

3. One ring to bring them all

Let's assume we have identified our topic and defined a search strategy, for example:

```
(hematology[mh] OR hematologic diseases[mh]) AND (adolescent*[TiAb] OR teenager*[TiAb] OR "young adult"[TiAb] OR "young adults"[TiAb] OR aya[TiAb]) AND (((Share*[TiAb] OR sharing[TiAb] OR informed[TiAb] OR collaborat*[TiAb] ) AND (decision*[TiAb] OR deciding[TiAb] OR choice*[TiAb] OR care*[TiAb])) OR ((patient*[TiAb]) AND (preference*[TiAb] OR view*[TiAb] OR involvement[TiAb] OR decision[TiAb] OR "decision making"[TiAb] OR attitude*[TiAb] OR participation[TiAb])))
```

(That is the way to tell Pubmed we're interested into shared decision making for hematological young patients)

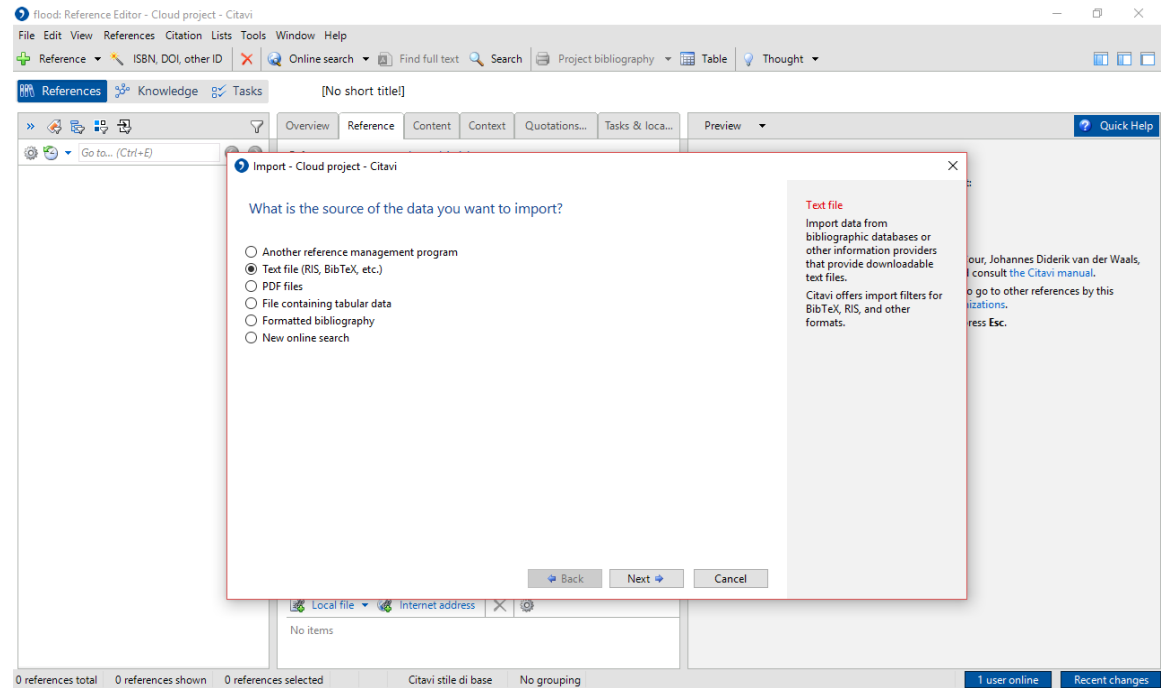
We have 192 results, and replicating the same query to Web of Science and Scopus, we reach 656.

Question: if we are extremely fast clickers (and lucky enough), how long will it take to download them?

Five hours and a half (without toilet breaks)

3. One ring to bring them all

Export your references from your database(s) and import the BibTeX file(s) in Citavi (or any other reference management software)



3. One ring to bring them all

Add the imported references to your project (and then perform an automatic duplicate check).

The screenshot shows the Citavi Reference Editor interface. A dialog box titled "Import" is open, displaying the following information:

- File import: savedrecs.bib (BibTeX)
- 170 references
- 170 / 170 (checked)
- 170 references imported successfully.

The dialog contains a table of imported references:

| <input checked="" type="checkbox"/> | Author or editor | Year | Title | Duplicate |
|-------------------------------------|-------------------------------|------|--|-----------|
| <input checked="" type="checkbox"/> | Schreuders, Elisabeth; Klappw | 2018 | Friend versus foe: Neural correlates of prosocial decisio | |
| <input checked="" type="checkbox"/> | Clement, S. C.; Kremer, L. C. | 2018 | Balancing the benefits and harms of thyroid cancer surv | |
| <input checked="" type="checkbox"/> | Simons, M. P.; Smietanski, M | 2018 | International guidelines for groin hernia management | |
| <input checked="" type="checkbox"/> | Cavazos-Rehg, Patricia A.; Kr | 2018 | Exposure to and Content of Marijuana Product Reviews | |
| <input checked="" type="checkbox"/> | Santamaria-Garcia, Hernand | 2018 | The interplay between sharing behavior and beliefs abo | |
| <input checked="" type="checkbox"/> | Stahmer, Aubyn C.; Suhrhein | 2018 | Exploring multi-level system factors facilitating educato | |
| <input checked="" type="checkbox"/> | Silva, Cristina; Almeida-Sant | | Antineoplastic Agents and (In)fertility: Informing Patient | |
| <input checked="" type="checkbox"/> | Aronson, Paul L.; Shapiro, Eu | 2018 | Shared Decision-Making with Parents of Acutely Ill Chil | |

Buttons at the bottom of the dialog: "Import again", "Add to project" (highlighted), and "Close".

On the right side of the dialog, there is an "Import" section with instructions: "To add all displayed references to your project, click **Add to project**. To add specific references only, first select the corresponding checkboxes and then click **Add to project**."

The background interface shows the "References" tab with a search bar and a "Quick Help" button.

3. One ring to bring them all

Select all your references and click “find full text”: the program will download (almost) every article that is available through University’s subscriptions.

In the meanwhile, you can have a coffee ;)

I used Citavi for these examples, keep in mind that Zotero is FOSS and works just as well.

The screenshot shows the Citavi Reference Editor interface. The window title is "flood: Reference Editor - Cloud project - Citavi". The menu bar includes File, Edit, View, References, Citation, Lists, Tools, Window, and Help. The toolbar contains icons for Reference, ISBN, DOI, other ID, Online search, Find full text, Search, Project bibliography, Table, Cite, and Thought. The main area is divided into three panes: References, Knowledge, and Tasks. The References pane shows a list of references, with the following one selected:

- Alden, Dana L.; Merz, Miwa Y.; Akashi, Jun
Young Adult Preferences for Physician Decision-Making Style in Japan and the United States
2012 – Journal Article

The Reference pane shows the details for the selected reference:

- Reference type: Journal Article
- Author: Alden, Dana L.; Merz, Miwa Y.; Akashi, Jun
- Title: Young Adult Preferences for Physician Decision-Making Style in Japan and the United States
- Subtitle:
- Title supplement:
- Collaborators:
- Periodical: ASIA-PACIFIC JOURNAL OF PUBLIC HEALTH
- Volume: 24
- Year: 2012
- Issue number: 1
- Page range: 173–184
- Article number:
- Online Address:
- Online since:
- Access date:
- Publisher:
- Database:
- DOI name: 10.1177/1010539510365098
- Notes: JAN
- Source: BibTeX
- Local file: [Web] doi.org/10.1177/1010539510365098

The Preview pane shows the author information and instructions for entering names:

Author
When entering a name, please use the following format:
Last name, then a comma, then the first name(s):
Garcia Lorca, Federico
Separate multiple names with a semicolon:
Plaget, Jean; Shaw, George Bernard
Things become complicated with names like: Paul La Cour, Johannes Didenik van der Waals, Catherine II. Use the input form (F9 or click Author) and consult the Citavi manual.
You can right-click the name of an author to edit it or to go to other references by this person. You can also open the list of persons and organizations.
To reject a suggestion from the list of existing names, press Esc.

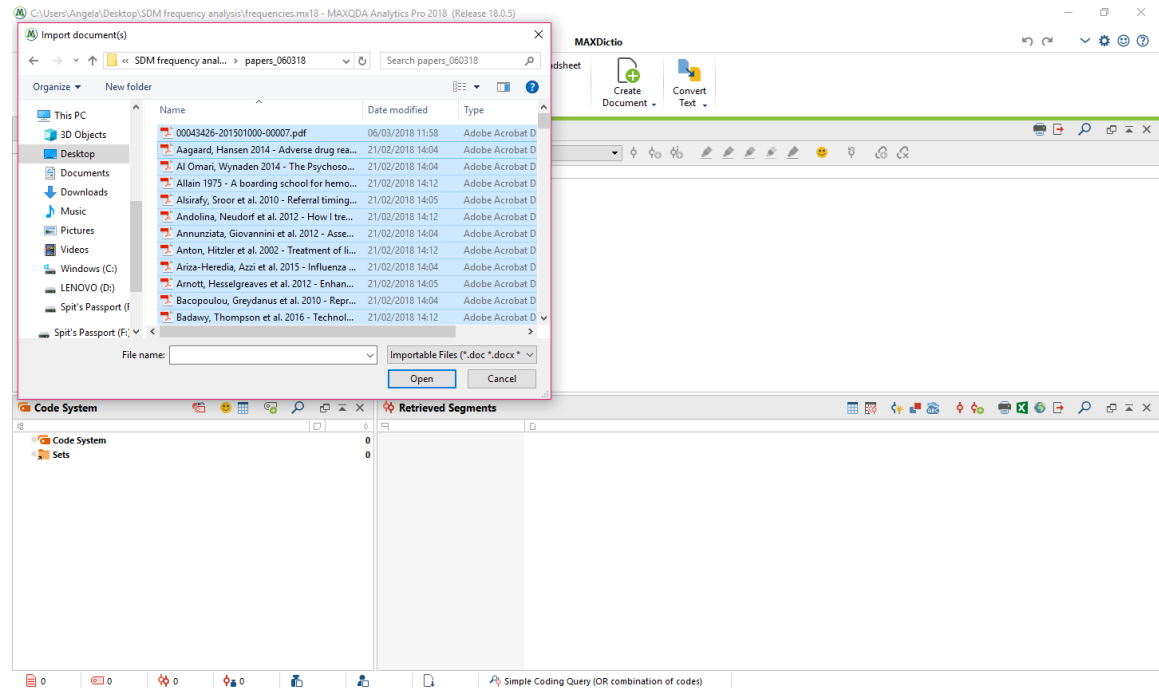
The status bar at the bottom shows: Cancel selection | 170 references total | 170 references shown | 170 references selected | Citavi Default Style | No grouping | 1 user online | Recent changes

4. One ring to 'read' them all

(yup this is made up)

4. One ring to 'read' them all

Import all the pdfs in MaxQDA Analytics pro.



4. One ring to 'read' them all

Activate them clicking on the gray dot close to the main folder.

NB: the program analyses only the active documents, so all these techniques are applicable both to the entire corpus, to a single paper, and every subset in between (e.g. only papers published in 2020)

The screenshot displays the MAXQDA Analytics Pro 2018 (Release 18.0.5) interface. The main window is titled 'Document System' and shows a list of documents under the 'Documents' folder. Each document has a small gray dot next to it, indicating its status. The 'Documents' folder is expanded, showing a list of documents with their titles and a '0' in the right column. The 'Document Browser' window is open, displaying a document titled 'Development and validation of a measure of disease-specific quality of life in young children with haemophilia'. The document content includes the title, authors (M. MANCO-JOHNSON, G. MORRISSEY-HARDING, B. EDELMAN-LEWIS, J. G. OSTERT and P. LARSON), and a summary. The 'Code System' window is also visible, showing a list of codes and sets. The 'Retrieved Segments' window is empty. The bottom status bar shows 'Simple Coding Query (OR combination of codes)'.

4. One ring to 'read' them all

Word frequency analysis: as a first passage, as we have done with Voyant Tools, we can see the frequency of single words.

In the same way, we can evaluate the frequency of groups of words. The results can be global or grouped per paper.

The screenshot displays the MAXQDA Analytics Pro 2018 interface. The main window shows a list of documents on the left and a 'Word frequencies' window in the foreground. The 'Word frequencies' window displays a table of the most frequent words across 153 documents, totaling 576648 words. The table includes columns for Word, Word length, Frequency, %, Rank, Documents, and Documents %.

| Word | Word length | Frequency | % | Rank | Documents | Documents % |
|------------|-------------|-----------|------|------|-----------|-------------|
| patient | 7 | 7.776 | 1,35 | 1 | 152 | 99,35 |
| that | 4 | 6.590 | 1,14 | 2 | 153 | 100,00 |
| cancer | 6 | 6.049 | 1,05 | 3 | 132 | 86,27 |
| have | 4 | 5.258 | 0,91 | 4 | 153 | 100,00 |
| care | 4 | 4.918 | 0,85 | 5 | 148 | 96,73 |
| study | 5 | 4.071 | 0,71 | 6 | 151 | 98,69 |
| child | 5 | 3.740 | 0,65 | 7 | 126 | 82,35 |
| health | 6 | 3.370 | 0,58 | 8 | 149 | 97,39 |
| from | 4 | 3.181 | 0,55 | 9 | 153 | 100,00 |
| treatment | 9 | 3.099 | 0,54 | 10 | 148 | 96,73 |
| much | 4 | 2.961 | 0,51 | 11 | 153 | 100,00 |
| parent | 6 | 2.686 | 0,47 | 12 | 103 | 67,32 |
| adolescent | 10 | 2.685 | 0,47 | 13 | 108 | 70,59 |
| use | 3 | 2.630 | 0,46 | 14 | 152 | 99,35 |
| about | 5 | 2.478 | 0,43 | 15 | 147 | 96,08 |
| they | 4 | 2.230 | 0,39 | 16 | 149 | 97,39 |
| decision | 8 | 2.071 | 0,36 | 17 | 106 | 69,28 |
| research | 8 | 2.050 | 0,36 | 18 | 146 | 95,42 |

4. One ring to ‘read’ them all

For example, now we know that 81 papers (52,94%) mention “quality of life”, that was not included in the query. This passage is useful to have an overall view on the corpus and to define further exploration strategies.

| Word combination | Frequency | % | Rank | Documents | Documents % |
|-----------------------|-----------|------|------|-----------|-------------|
| sickle cell disease | 403 | 0,17 | 1 | 33 | 21,57 |
| quality of life | 393 | 0,16 | 2 | 81 | 52,94 |
| end of life | 225 | 0,09 | 3 | 39 | 25,49 |
| in decision make | 197 | 0,08 | 4 | 27 | 17,65 |
| of pediatric oncology | 172 | 0,07 | 5 | 33 | 21,57 |
| to participate in | 161 | 0,07 | 6 | 53 | 34,64 |
| much likely to | 149 | 0,06 | 7 | 62 | 40,52 |
| pediatr blood cancer | 131 | 0,05 | 8 | 30 | 19,61 |
| parent of child | 130 | 0,05 | 9 | 34 | 22,22 |
| health care provider | 120 | 0,05 | 10 | 35 | 22,88 |

4. One ring to 'read' them all

Dictionaries: in this case, we are interested in understanding what the literature says about certain topics relevant for the research question, namely autonomy, responsibility, decision, guidance, impact, care, patient, physician, nurse, family, disease and shared decision making.

We can define a dictionary of synonyms and use them to run a paper and category based frequency analysis.

The screenshot displays the MAXQDA Analytics Pro interface. The main window shows a list of documents under the 'Documents' tab, with a 'Document Browser' window open for a specific document. A 'Dictionary' configuration window is overlaid on top, showing a tree view of categories and a table of search items.

The 'Dictionary' window shows the following structure:

- Categories
 - Root
 - autonomy
 - responsibility
 - decision
 - guidance
 - impact
 - care
 - patient
 - physician
 - nurse
 - family
 - disease
 - SDM

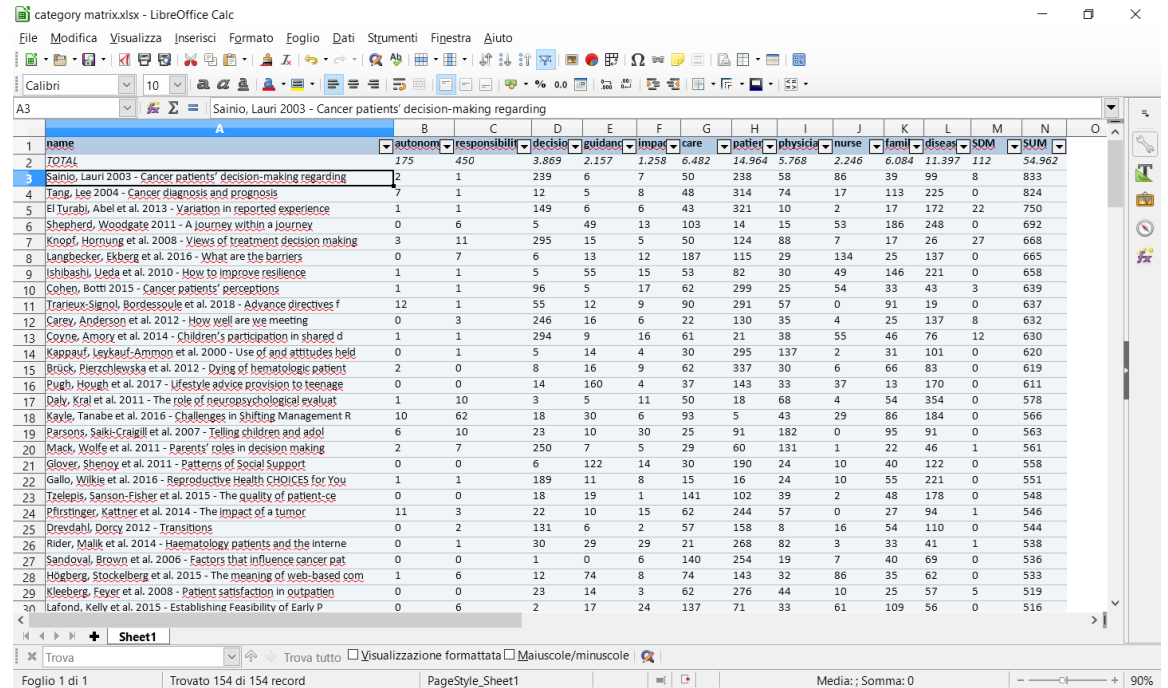
The 'Search items' table is as follows:

| Search items | Whole word | Case sensitivity | Starting letters |
|------------------|--------------------------|--------------------------|--------------------------|
| responsibility | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| responsibilities | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| decision | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| responsible of | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| responsible for | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| impact | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| guardian | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| guardians | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| guardianship | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| tutor | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| tutors | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| tutorship | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| delegate | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| delegates | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| in charge of | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| supervision | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

4. One ring to 'read' them all

The resulting excel spreadsheet offers almost endless possibilities: for instance, we can identify at a glance the papers that most likely will be more important for a certain topic alone (e.g: autonomy) or for any combination of different topics (e.g: autonomy and family).

Moreover, from the top row we have some insights on the corpus in general, knowing that for example the concept of responsibility is more debated than autonomy, that the role of physicians is less debated than the role of families but more than the role of nurses, and so on.



category matrix.xlsx - LibreOffice Calc

File Modifica Visualizza Inserisci Formato Foglio Dati Strumenti Finestra Aiuto

Calibri 10

A3 Sainio, Lauri 2003 - Cancer patients' decision-making regarding

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|---|---------|--------------|---------|---------|-------|-------|--------|---------|-------|-------|--------|-----|--------|---|
| 1 | name | autonom | responsibili | decisio | guidanz | impac | care | patier | physica | nurse | famil | diseas | SOM | SUM | |
| 2 | TOTAL | 175 | 450 | 3.869 | 2.157 | 1.258 | 6.482 | 14.964 | 5.768 | 2.246 | 6.084 | 11.397 | 112 | 54.962 | |
| 3 | Sainio, Lauri 2003 - Cancer patients' decision-making regarding | 2 | 1 | 239 | 6 | 7 | 50 | 238 | 58 | 86 | 39 | 99 | 8 | 833 | |
| 4 | Tang, Lee 2004 - Cancer diagnosis and prognosis | 7 | 1 | 12 | 5 | 8 | 48 | 314 | 74 | 17 | 113 | 225 | 0 | 824 | |
| 5 | El Turabi, Abel et al. 2013 - Variation in reported experience | 1 | 1 | 149 | 6 | 6 | 43 | 321 | 10 | 2 | 17 | 172 | 22 | 750 | |
| 6 | Shepherd, Woodgate 2011 - A journey within a journey | 0 | 6 | 5 | 49 | 13 | 103 | 14 | 15 | 53 | 186 | 248 | 0 | 692 | |
| 7 | Knopf, Hornung et al. 2008 - Views of treatment decision making | 3 | 11 | 295 | 15 | 5 | 50 | 124 | 88 | 7 | 17 | 26 | 27 | 668 | |
| 8 | Langbecker, Ekberg et al. 2016 - What are the barriers | 0 | 7 | 6 | 13 | 12 | 187 | 115 | 29 | 134 | 25 | 137 | 0 | 665 | |
| 9 | Ishibashi, Ueda et al. 2010 - How to improve resilience | 1 | 1 | 5 | 55 | 15 | 53 | 82 | 30 | 49 | 146 | 221 | 0 | 658 | |
| 10 | Cohen, Boti 2015 - Cancer patients' perceptions | 1 | 1 | 96 | 5 | 17 | 62 | 299 | 25 | 54 | 33 | 43 | 3 | 639 | |
| 11 | Tranieux-Signol, Bordessoule et al. 2018 - Advance directives f | 12 | 1 | 55 | 12 | 9 | 90 | 291 | 57 | 0 | 91 | 19 | 0 | 637 | |
| 12 | Carey, Anderson et al. 2012 - How well are we meeting | 0 | 3 | 246 | 16 | 6 | 22 | 130 | 35 | 4 | 25 | 137 | 8 | 632 | |
| 13 | Coyne, Amory et al. 2014 - Children's participation in shared d | 1 | 1 | 294 | 9 | 16 | 61 | 21 | 38 | 55 | 46 | 76 | 12 | 630 | |
| 14 | Kappauf, Leykauf-Ammon et al. 2000 - Use of and attitudes held | 0 | 1 | 5 | 14 | 4 | 30 | 295 | 137 | 2 | 31 | 101 | 0 | 620 | |
| 15 | Brück, Pierzchlewska et al. 2012 - Dying of hematologic patient | 2 | 0 | 8 | 16 | 9 | 62 | 337 | 30 | 6 | 66 | 83 | 0 | 619 | |
| 16 | Pugh, Hough et al. 2017 - Lifestyle advice provision to teenage | 0 | 0 | 14 | 160 | 4 | 37 | 143 | 33 | 37 | 13 | 170 | 0 | 611 | |
| 17 | Daly, Kral et al. 2011 - The role of neuropsychological evaluat | 1 | 10 | 3 | 5 | 11 | 50 | 18 | 68 | 4 | 54 | 354 | 0 | 578 | |
| 18 | Kayle, Tanabe et al. 2016 - Challenges in Shifting Management R | 10 | 62 | 18 | 30 | 6 | 93 | 5 | 43 | 29 | 86 | 184 | 0 | 566 | |
| 19 | Parsons, Saiki-Craigill et al. 2007 - Telling children and adol | 6 | 10 | 23 | 10 | 30 | 25 | 91 | 182 | 0 | 95 | 91 | 0 | 563 | |
| 20 | Mack, Wolfe et al. 2011 - Parents' roles in decision making | 2 | 7 | 250 | 7 | 5 | 29 | 60 | 131 | 1 | 22 | 46 | 1 | 561 | |
| 21 | Glover, Shenoy et al. 2011 - Patterns of Social Support | 0 | 0 | 6 | 122 | 14 | 30 | 190 | 24 | 10 | 40 | 122 | 0 | 558 | |
| 22 | Gallo, Wilkie et al. 2016 - Reproductive Health CHOICES for You | 1 | 1 | 189 | 11 | 8 | 15 | 16 | 24 | 10 | 55 | 221 | 0 | 551 | |
| 23 | Tzelepis, Sanson-Fisher et al. 2015 - The quality of patient-ce | 0 | 0 | 18 | 19 | 1 | 141 | 102 | 39 | 2 | 48 | 178 | 0 | 548 | |
| 24 | Pfirsinger, Kattner et al. 2014 - The impact of a tumor | 11 | 3 | 22 | 10 | 15 | 62 | 244 | 57 | 0 | 27 | 94 | 1 | 546 | |
| 25 | Drevdahl, Dorcy 2012 - Transitions | 0 | 2 | 131 | 6 | 2 | 57 | 158 | 8 | 16 | 54 | 110 | 0 | 544 | |
| 26 | Rider, Malik et al. 2014 - Haematology patients and the interne | 0 | 1 | 30 | 29 | 29 | 21 | 268 | 82 | 3 | 33 | 41 | 1 | 538 | |
| 27 | Sandoval, Brown et al. 2006 - Factors that influence cancer pat | 0 | 0 | 1 | 0 | 6 | 140 | 254 | 19 | 7 | 40 | 69 | 0 | 536 | |
| 28 | Högberg, Stockelberg et al. 2015 - The meaning of web-based com | 1 | 6 | 12 | 74 | 8 | 74 | 143 | 32 | 86 | 35 | 62 | 0 | 533 | |
| 29 | Kleeberg, Feyrer et al. 2008 - Patient satisfaction in outpatient | 0 | 0 | 23 | 14 | 3 | 62 | 276 | 44 | 10 | 25 | 57 | 5 | 519 | |
| 30 | Lafond, Kelly et al. 2015 - Establishing Feasibility of Early P | 0 | 6 | 2 | 17 | 24 | 137 | 71 | 33 | 61 | 109 | 56 | 0 | 516 | |

Foglio 1 di 1 Trovato 154 di 154 record PageStyle_Sheet1 Media: ; Somma: 0 90%

4. One ring to 'read' them all

As a last treat, MaxQDA can autocode documents using our dictionaries: we can build a set of “subcorpus” containing all the sentences that contain a specific set of keywords.

This is useful if for instance we want to explore what are the most commonly debated concepts in the sentences concerning autonomy, or patients, or any other category.

The screenshot displays the MAXQDA Analytics Pro 2018 interface. The top menu bar includes Home, Import, Codes, Variables, Analysis, Mixed Methods, Visual Tools, Reports, Stats, and MAXDictio. The MAXDictio menu is open, showing options like Word Frequencies, Stop List, Go List, Keyword-in-context, Word Combinations, Interactive Word Tree, Quantitative Content Analysis, Dictionary, Category Matrix Browser, Autocode with Dictionary, and Language Options. The 'Autocode with Dictionary' option is selected, opening a dialog box. The dialog box has a 'Code' section with a dropdown set to 'autonomy' and a 'With weight' field set to '0'. Below this, there are sections for 'PDF documents' and 'Text and table documents', each with radio buttons for 'Only search string' and 'Context', and dropdowns for 'Sentences before' and 'Words after'. The 'Autocode' button is highlighted. In the background, the 'Document System' pane shows a list of documents, and the 'Code System' pane shows a tree structure with 'MAXDictio' expanded to show various codes like 'autonomy', 'responsibility', 'decision', etc. The 'Retrieved' pane shows a list of documents with counts.

The nose problem

The value of these data is actually pretty low: from a frequency point of view the sentences “my nose is nice” and “my nose is not nice” are identical, because they both mention “nose” and “nice”.

These are not final results: they are powerful hints to plan the next steps.

For more details about this stuff: <https://doi.org/10.1016/j.heliyon.2020.e04426>

**5. Here it gets very nerdy,
or: The Mighty Topic Tracker**

What we can do so far:

- Perform and expand queries by text mining titles and abstracts (DB search engines + Voyant Tools)
- Download full texts using a reference management software
- Very basic full text analysis in MaxQDA

Hint: there's more in a db entry than meets the eye! Let's explore further this material with some Python-based NLP.

6. Here it gets very nerdy

Jupyter is a project and community whose goal is to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages".

Project Jupyter's name is a reference to the three core programming languages supported by Jupyter, which are Julia, Python and R, and also a homage to Galileo's notebooks recording the discovery of the moons of Jupiter.

Project Jupyter has developed and supported the interactive computing products Jupyter Notebook, JupyterHub, and JupyterLab.



6. Here it gets very nerdy

Why Python?

- it's high level (strong abstraction, little need to know what happens under the code)
- It's general purpose (you can write almost anything, ranging from a robot to a web application)
- It's easy to read and to learn (when compared with other programming languages)
- It's logical and tidy
- It's widely used and there's a ton of libraries to extend its core
- It's cross-platform (i.e. you can run Python code on Windows, Linux and MacOS)



6. Here it gets very nerdy

Anaconda

A distribution of Python and R aimed to scientific programming that simplifies a lot installation, package management, environment management.



ANACONDA®

6. Here it gets very nerdy

The screenshot displays the Anaconda Navigator application window. The interface is divided into several sections:

- Left Sidebar:** Contains navigation options: Home, Environments (highlighted with a blue arrow), Learning, and Community. Below these are promotional cards for ANACONDA NUCLEUS, Discover premium data science content, Documentation, and Anaconda Blog, along with social media icons for Twitter, YouTube, and GitHub.
- Environments Panel:** A list of environments including 'base (root)', 'PubliCo scorer v1', 'TelegramHistory', 'TopicTracker' (highlighted with a blue arrow), 'dsi tap', 'igpattern', 'rstudio', 'rstudio-', and 'rstudio_1'. A search bar is located at the top of this panel.
- Packages Panel:** A table of installed packages with columns for Name, Description, and Version. A blue arrow points from the 'TopicTracker' environment to this panel. The table lists various Python packages such as appdirs, argon2-cffi, async-generator, attrs, backcall, blas, bleach, blis, bokeh, brotli, ca-certificates, catalogue, certifi, and cffi.

External labels with blue arrows point to the 'Environments' and 'Packages' sections of the interface.

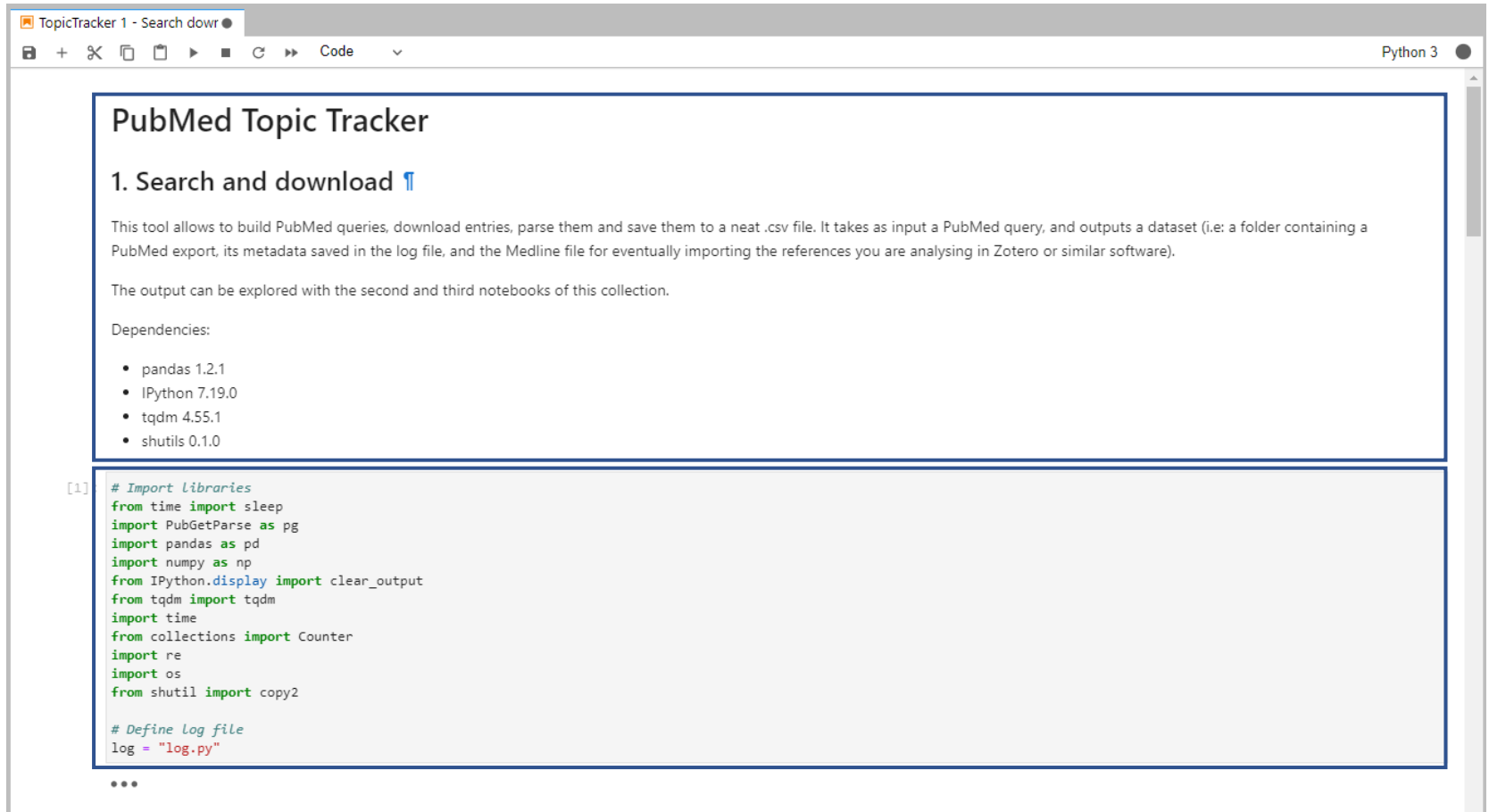
| Name | Description | Version |
|-----------------|---|-----------|
| appdirs | A small python module for determining appropriate platform-specific dirs. | 1.4.4 |
| argon2-cffi | The secure argon2 password hashing algorithm. | 20.1.0 |
| async-generator | | 1.10 |
| async_generator | Async generators and context managers for python 3.5+ | 1.10 |
| attrs | Attrs is the python package that will bring back the joy of writing classes by relieving you from the drudgery of implementing object protocols (aka dunder methods). | 20.3.0 |
| backcall | Specifications for callback functions passed in to an api | 0.2.0 |
| blas | | 1.0 |
| bleach | Easy, whitelist-based html-sanitizing tool | 3.2.1 |
| blis | | 0.4.1 |
| bokeh | Statistical and novel interactive html plots for python | 2.2.3 |
| brotli | Python bindings to the brotli compression library | 0.7.0 |
| ca-certificates | Certificates for use with other packages. | 2020.12.5 |
| catalogue | Super lightweight function registries for your library | 1.0.0 |
| certifi | Python package for providing mozilla's ca bundle. | 2020.12.5 |
| cffi | Foreign function interface for python calling c code. | 1.14.4 |

6. Here it gets very nerdy

Basic structure of a JupyterLab notebook

Comment blocks
In markdown

Code blocks
In this case it's Python



The screenshot shows a JupyterLab notebook interface. The browser title bar reads "TopicTracker 1 - Search down". The notebook content is as follows:

PubMed Topic Tracker

1. Search and download ¶

This tool allows to build PubMed queries, download entries, parse them and save them to a neat .csv file. It takes as input a PubMed query, and outputs a dataset (i.e: a folder containing a PubMed export, its metadata saved in the log file, and the Medline file for eventually importing the references you are analysing in Zotero or similar software).

The output can be explored with the second and third notebooks of this collection.

Dependencies:

- pandas 1.2.1
- IPython 7.19.0
- tqdm 4.55.1
- shutils 0.1.0

```
[1] # Import Libraries
from time import sleep
import PubGetParse as pg
import pandas as pd
import numpy as np
from IPython.display import clear_output
from tqdm import tqdm
import time
from collections import Counter
import re
import os
from shutil import copy2

# Define Log file
log = "log.py"
```

... (three dots)

6. Here it gets very nerdy

Why should I bother learning this stuff? I'm no programmer after all, right?

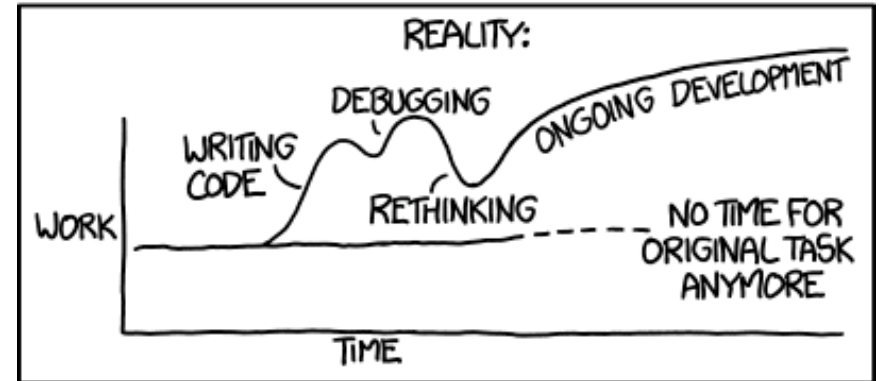
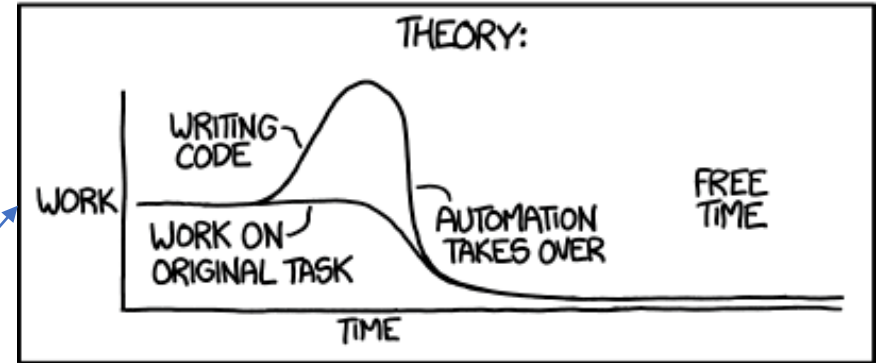
True. But we live in a world packed with data, and these are excellent tools to work on data. In Python you can do NLP, score a survey, map networks, autocode texts using very refined rules, and water the plants if need be. It's a skill worth acquiring.

Some cool examples I actually wrote and used for my research at the IBME:

- [Factiva parser and NLP analysis](#)
- [Telegram social listening](#)

In this case, the "theory" case applies (most of the times!)

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



6. Here it gets very nerdy

Finding TopicTracker
Easy. As all my software, it's available on my Zenodo repo under a CC-BY license. Just google "zenodo topictracker" and download the latest version (currently 1.2.0)

The screenshot shows the Zenodo repository page for 'TopicTracker: a Python pipeline to search, download and explore PubMed entries'. The page is dated February 8, 2021, and is categorized as 'Software' and 'Open Access'. The title is prominently displayed at the top. Below the title, the authors are listed as Giovanni Spitale and Nikola Biller-Andorno. A brief description states that TopicTracker is a Python pipeline intended to streamline and simplify the retrieval and exploration of large amounts of PubMed entries, divided into three Jupyter notebooks: 1. Search and download; 2. Content analyser; 3. Interactive data exploration. The page details three notebooks: the first for building and parsing PubMed queries, the second for analyzing trends over time, and the third for interactive exploration of datasets. It also lists dependencies and provides links to download toy datasets. On the right side, there are buttons for 'Edit', 'New version', and 'Communities' (scientometrics, University of Zurich). Statistics show 386 views and 31 downloads. The page is indexed in OpenAIRE. The publication date is February 8, 2021, with a DOI of 10.5281/zenodo.4707825. The keyword(s) are 'natural language processing, scientometrics, topic tracking, information extraction'.

zenodo Search Upload Communities g.jovanni.spitale@uzh.ch

February 8, 2021 Software Open Access Edit

TopicTracker: a Python pipeline to search, download and explore PubMed entries

Giovanni Spitale; Nikola Biller-Andorno

TopicTracker is a Python pipeline intended to streamline and simplify the retrieval and exploration of large amounts of PubMed entries. The software is divided into three Jupyter notebooks: 1. Search and download; 2. Content analyser; 3. Interactive data exploration.

The first notebook allows to build PubMed queries, download entries, parse them and save them to a .csv file. It takes as input a PubMed query, and outputs a dataset (i.e. a folder containing a PubMed export, its metadata saved in the log file, and the Medline file for eventually importing the references you are analysing in Zotero or similar software). The functions for searching, downloading and parsing are written in a different module in order to simplify adaptations for other projects if need be. The output of the first notebook can be explored with the second and third notebooks of this collection.

The second notebook allows to analyse the trends of entities over time. It takes as input a dataset (i.e. a folder containing a PubMed export generated with the first notebook of this collection, its metadata, and the Medline file) and it outputs a set of .csv files and .svg plots with the trends of keywords, MeSH terms, authors, journals, lemmas in Title/Abstract, amount of COI statements, lemma trends in COI statements. The .csv files can then be explored further with the third notebook of this collection.

The third notebook allows fully interactive exploration of the datasets preprocessed with the second notebook. You can select a dataset to work with, a set of entities to explore, and plot any entity or combination of entities.

Dependencies (and versions) are listed in every notebook. A couple of toy datasets are provided.

New in v 1.2:

- managed some more exceptions
- streamlined the NLP pipeline
- optimized details here and there
- included LARGE toy datasets for more fun in learning the software :)

To do in v1.3:

- understand why the PubMed APIs work so strangely with the PDAT tag
- manage exceptions (=empty files) in the tabs of notebook 3
- optimize the creation of the medline file in notebook 1

Communities: scientometrics, University of Zurich

386 views, 31 downloads

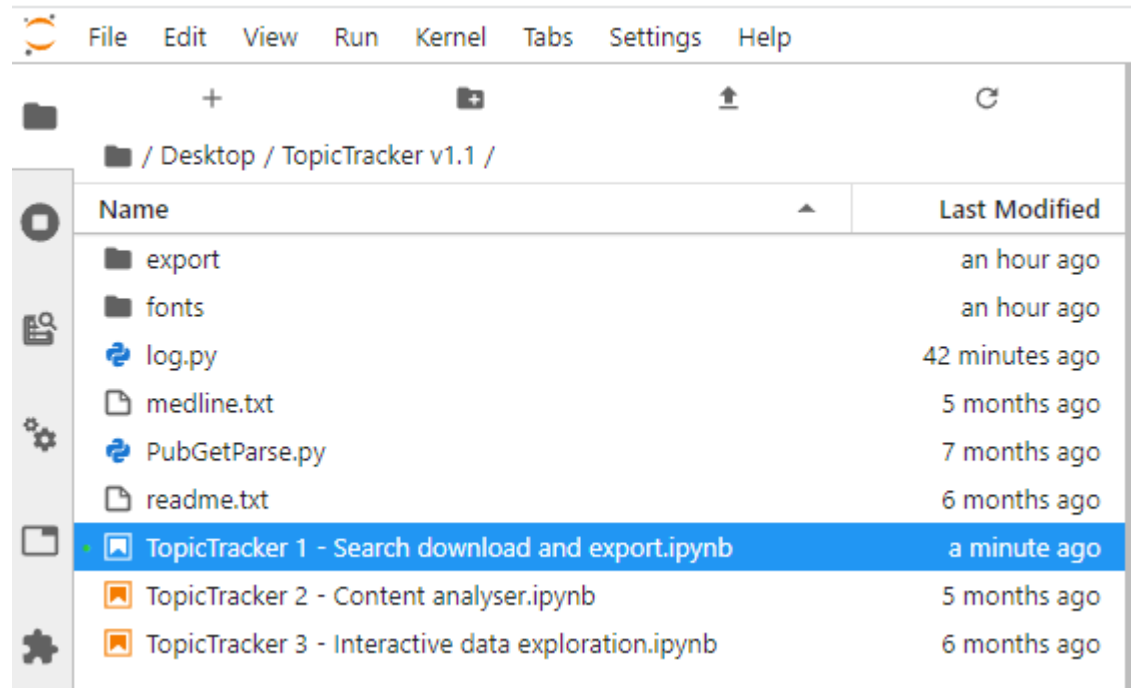
Indexed in OpenAIRE

Publication date: February 8, 2021
DOI: 10.5281/zenodo.4707825
Keyword(s): natural language processing, scientometrics, topic tracking, information extraction

Preview

6. Here it gets very nerdy

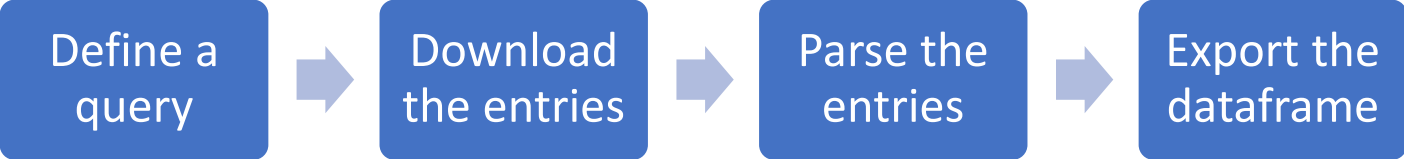
Running TopicTracker
Easy. Just install Anaconda, create a new environment, install the dependencies listed in the notebooks, unzip the download in your desktop, launch JupyterLab, navigate to the right folder and open the first notebook by double clicking on it.





6. Here it gets very nerdy

Notebook 1 – Search, download and export



6. Here it gets very nerdy

Notebook 1 – Search, download and export

This doesn't do anything super fancy – basically it runs queries, gets entries like this...

and transform them into a neat table like this one here:

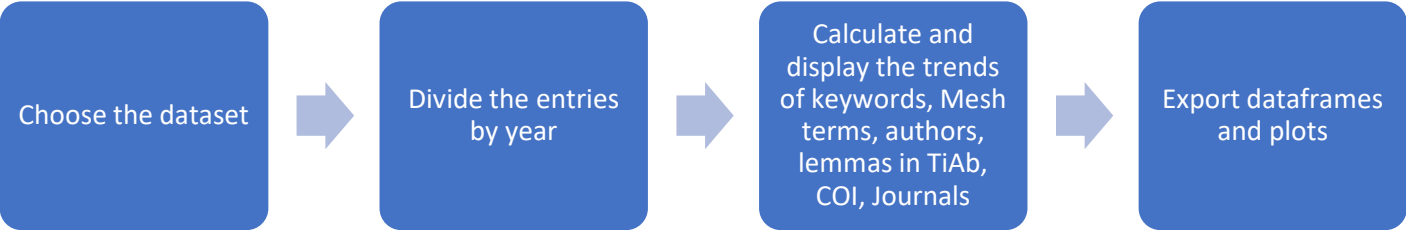
PMID- 34612823
 OWN - NLM
 STAT - Publisher
 LR - 20211006
 IS - 1929-0748 (Print)
 IS - 1929-0748 (Linking)
 DP - 2021 Sep 22
 TI - PubliCo. A protocol for a new risk and crisis communication platform to bridge the gap between policy makers and the public in the context of the COVID-19 crisis.
 LID - 10.2196/33653 [doi]
 AB - BACKGROUND: Since the end of 2019, COVID-19 has had a significant impact on citizens around the globe. As governments institute more restrictive measures, public adherence could decrease and discontent mount. Providing high-quality information and countering fake news is important. But we also need feedback loops so that government officials can refine preventive measures and communication strategies. Policy-makers need information - preferably based on real-time data - on the public's cognitive, emotional and behavioural reaction to public health messages and restrictive measures. PubliCo aims to foster effective and tailored risk and crisis communication as well as an assessment of the risks and benefits of prevention and control measures, as their effectiveness depends on public trust and cooperation. OBJECTIVE: Our project aims to develop a tool that helps tackle the COVID-19 infodemic, with a focus on enabling a nuanced and in-depth understanding of public perception. The project adopts a trans-disciplinary multi-stakeholder approach, including participatory citizen science. METHODS: We combine literature and media review and analysis and empirical research using mixed methods, including an online survey and diary-based research, both of which are ongoing and continuously updated. Building on real-time data and continuous data collection, our research results will be highly adaptable to the evolving situation. RESULTS: As of September 2021, two thirds of the tool we propose are up and running. Current development cycles focus on the analytics component, on user experience, and on interface refinements. We collected a total of 473 responses through PubliCo Survey, and diaries through PubliCo Diaries. CONCLUSIONS: Pilot data show that PubliCo is a promising and efficient concept for bidirectional risk and crisis communication in the context of public health crises; further data are needed to assess its function at a larger scale or in the context of an issue other than COVID-19. CLINICALTRIAL:
 FAU - Spitale, Giovanni
 AU - Spitale G
 AD - Institute of Biomedical Ethics and History of Medicine, University of Zurich, Winterthurerstrasse 30. Zurich. CH.

| | p_id | pid_type | year | journal | publisher | title | book_title | abstract | oabstract | authors | editors | language | meshterms | keywords | coi | grant | doi |
|---|----------|----------|------|--|-----------|---|------------|---|-----------|-----------------------------------|---------|----------|---|---|-----|-------|-------------------------------|
| 1 | 12349742 | Article | 2000 | AIDS weekly | NaN | Malaysia urges ASEAN to tackle AIDS crisis. | NaN | Urgent action is needed to fight the alarming ... | NaN | NaN | NaN | eng | ['*Acquired Immunodeficiency Syndrome', 'Asia'... | ['*Acquired Immunodeficiency Syndrome', 'Asia'... | NaN | NaN | NaN |
| 2 | 10795397 | Article | 2000 | Occupational medicine (Oxford, England) | NaN | A South-East Asian perspective. | NaN | In order to discuss the subject of occupatio... | NaN | Koh D, Chia SE, Jeyaratnam J | NaN | eng | ['Asia, Southeastern/epidemiology', 'Delivery ... | NaN | NaN | NaN | 10.1093/occmed/50.1.64 |
| 3 | 14602265 | Article | 2003 | The Annals of thoracic surgery | NaN | Peritoneal dialysis after surgery for congenit... | NaN | BACKGROUND: We determined the risk factors for... | NaN | Chan KL, Ip P, Chiu CS, Cheung YF | NaN | eng | ['Acute Kidney Injury/etiology/'mortality/'the... | NaN | NaN | NaN | 10.1016/s0003-4975(03)01026-9 |
| 4 | 14600111 | Article | 2003 | Journal of epidemiology and community health | NaN | Monitoring community responses to the SARS epi... | NaN | STUDY OBJECTIVE: To report the evolution in pe... | NaN | Lau JT, Yang X, Tsui H, Kim JH | NaN | eng | ['Adolescent', 'Adult', 'Attitude to Health', ... | NaN | NaN | NaN | 10.1136/jech.57.11.864 |



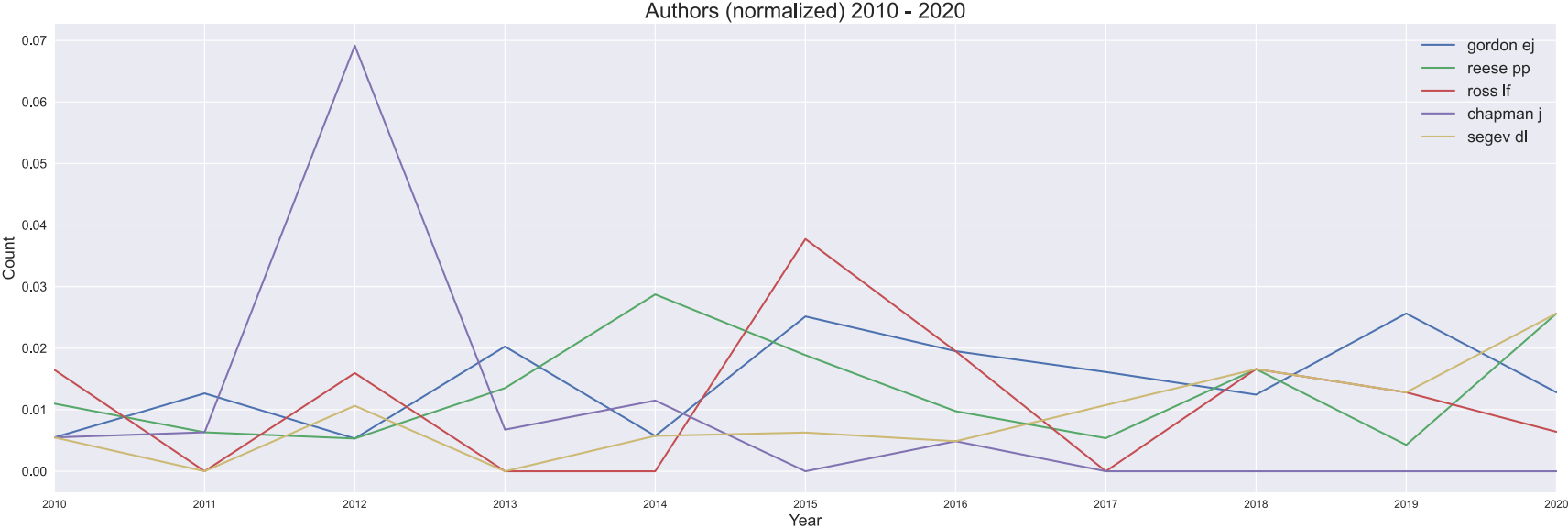
6. Here it gets very nerdy

Notebook 2 – Content analysis



6. Here it gets very nerdy

Notebook 2 – Content analysis (organ transplantation and ethics – authors, normalized)



6. Here it gets very nerdy

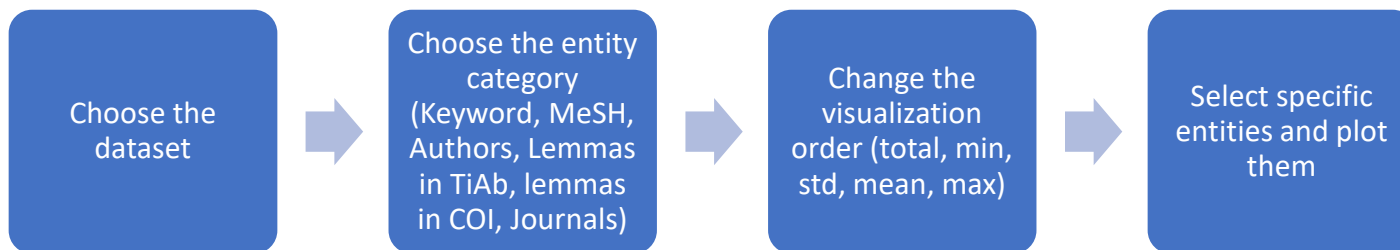
Notebook 2 – Content analysis (organ transplantation and ethics – journals, normalized)

| journal | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | total | min | std | mean | max |
|--|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| american journal of transplantation | 0,06 | 0,03 | 0,05 | 0,03 | 0,06 | 0,12 | 0,16 | 0,20 | 0,12 | 0,14 | 0,23 | 0,97 | 0,03 | 0,06 | 0,10 | 0,20 |
| transplantation | 0,03 | 0,05 | 0,05 | 0,03 | 0,06 | 0,06 | 0,05 | 0,06 | 0,03 | 0,03 | 0,04 | 0,47 | 0,03 | 0,01 | 0,05 | 0,06 |
| journal of medical ethics | 0,02 | 0,03 | 0,02 | 0,04 | 0,06 | 0,06 | 0,05 | 0,04 | 0,02 | 0,03 | 0,04 | 0,39 | 0,02 | 0,02 | 0,04 | 0,06 |
| transplantation proceedings | 0,05 | 0,09 | 0,04 | 0,04 | 0,03 | 0,04 | 0,00 | 0,02 | 0,02 | 0,04 | 0,04 | 0,37 | 0,00 | 0,02 | 0,04 | 0,09 |
| the american journal of bioethics | 0,07 | 0,13 | 0,04 | 0,00 | 0,03 | 0,04 | 0,01 | 0,01 | 0,01 | 0,00 | 0,01 | 0,33 | 0,00 | 0,04 | 0,03 | 0,13 |
| bmj open bioethics | 0,00 0,01 | 0,00 0,01 | 0,00 0,01 | 0,01 0,01 | 0,00 0,02 | 0,05 0,02 | 0,02 0,01 | 0,05 0,01 | 0,02 0,04 | 0,06 0,01 | 0,05 0,01 | 0,21 0,16 | 0,00 0,01 | 0,02 0,01 | 0,02 0,02 | 0,06 0,04 |
| liver transplantation | 0,02 | 0,03 | 0,01 | 0,03 | 0,01 | 0,03 | 0,00 | 0,02 | 0,01 | 0,00 | 0,00 | 0,16 | 0,00 | 0,01 | 0,02 | 0,03 |
| cambridge quarterly of healthcare ethics | 0,01 | 0,03 | 0,02 | 0,05 | 0,01 | 0,00 | 0,01 | 0,00 | 0,02 | 0,01 | 0,00 | 0,16 | 0,00 | 0,02 | 0,02 | 0,05 |
| progress in transplantation | 0,03 | 0,01 | 0,01 | 0,01 | 0,02 | 0,00 | 0,01 | 0,02 | 0,01 | 0,02 | 0,01 | 0,14 | 0,00 | 0,01 | 0,01 | 0,03 |



6. Here it gets very nerdy

Notebook 3 – Interactive data exploration





6. Here it gets very nerdy

Notebook 3 – Interactive data exploration (Bioethics subset 2015 - 2021)

Datasets available:

| | dirname | year0 | year1 | query | paper_count | querydatetime |
|---|-----------------|-------|-------|---|-------------|---------------------|
| 1 | 20210329-104455 | 2010 | 2020 | organ transplantation AND ethics | 2034 | 2021.03.29-09:54:03 |
| 2 | 20210329-112955 | 1988 | 2021 | "grappa"[All Fields] AND alcohol[All Fields] NOT psoriasis[All Fields] | 29 | 2021.03.29-11:29:11 |
| 3 | 20210415-224439 | 2015 | 2021 | bioethics[sb] | 74746 | 2021.04.15-09:18:15 |
| 4 | 20210427-143313 | 2000 | 2005 | (((((("risk"[MeSH Terms] OR "risk"[All Fields]) AND "crisis"[All Fields] AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Terms] OR "communication"[All Fields] OR "communications"[All Fields] OR "communicators"[All Fields])))) OR "crisis"[All Fields]) AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Ter... | 20 | 2021.04.27-14:32:59 |
| 5 | 20210531-150441 | 2010 | 2021 | (((((("risk"[MeSH Terms] OR "risk"[All Fields]) AND "crisis"[All Fields] AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Terms] OR "communication"[All Fields] OR "communications"[All Fields] OR "communicators"[All Fields])))) OR "crisis"[All Fields]) AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Ter... | 232 | 2021.05.31-15:02:08 |
| 6 | 20210531-154820 | 2000 | 2021 | (((((("risk"[MeSH Terms] OR "risk"[All Fields]) AND "crisis"[All Fields] AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Terms] OR "communication"[All Fields] OR "communications"[All Fields] OR "communicators"[All Fields])))) OR "crisis"[All Fields]) AND ("communicate"[All Fields] OR "communicated"[All Fields] OR "communicates"[All Fields] OR "communicating"[All Fields] OR "communication"[MeSH Ter... | 273 | 2021.05.31-15:45:22 |

Here you can choose the different datasets (downloaded and parsed with notebook 1 and analyzed with notebook 2).

The input field is dynamic, so you can start typing and it will autocomplete suggesting existing datasets.

...

Dataset:

Select

Using dataset: "20210415-224439"

Time interval: 2015 - 2021
Keywords: bioethics[sb]
74746 entries

Dataset ready for analysis.

6. Here it gets very nerdy

Notebook 3 – Interactive data exploration (Bioethics subset 2015 - 2021)

The tabs represent the categories of entities you can play with.

‘Sort by’ lets you specify how to sort the data (as the dataset is big, you will only see the top 20)

| Keyword | MeSH | Authors | Lemmas in TiAb | Lemmas in COI | Journals | | | | | | | | |
|----------|-----------------------|---------|----------------|---------------|----------|---------|---------|-------|------|-----|-------|--------|-----|
| Sort by: | total | | | | | | | | | | | | |
| keywords | df_2015 | df_2016 | df_2017 | df_2018 | df_2019 | df_2020 | df_2021 | total | min | std | mean | max | |
| 1 | ethics | 491 | 309 | 331 | 279 | 330 | 443 | 257 | 2440 | 257 | 86.26 | 348.57 | 491 |
| 2 | informed consent | 151 | 114 | 117 | 95 | 114 | 136 | 80 | 807 | 80 | 23.69 | 115.29 | 151 |
| 3 | bioethics | 91 | 86 | 74 | 55 | 73 | 137 | 74 | 590 | 55 | 25.90 | 84.29 | 137 |
| 4 | palliative care | 94 | 56 | 71 | 72 | 78 | 94 | 40 | 505 | 40 | 19.51 | 72.14 | 94 |
| 5 | autonomy | 91 | 75 | 63 | 51 | 61 | 69 | 46 | 456 | 46 | 15.10 | 65.14 | 91 |
| 6 | research ethics | 109 | 56 | 62 | 48 | 42 | 90 | 46 | 453 | 42 | 25.26 | 64.71 | 109 |
| 7 | qualitative research | 56 | 56 | 55 | 45 | 66 | 51 | 29 | 358 | 29 | 11.63 | 51.14 | 66 |
| 8 | abortion | 68 | 38 | 57 | 48 | 47 | 69 | 18 | 345 | 18 | 17.85 | 49.29 | 69 |
| 9 | nursing | 43 | 46 | 40 | 48 | 70 | 51 | 22 | 320 | 22 | 14.29 | 45.71 | 70 |
| 10 | medical ethics | 48 | 48 | 61 | 30 | 53 | 55 | 19 | 314 | 19 | 14.94 | 44.86 | 61 |
| 11 | communication | 50 | 42 | 50 | 31 | 39 | 42 | 20 | 274 | 20 | 10.68 | 39.14 | 50 |
| 12 | advance care planning | 34 | 24 | 41 | 40 | 51 | 44 | 38 | 272 | 24 | 8.41 | 38.86 | 51 |
| 13 | professionalism | 40 | 37 | 49 | 33 | 53 | 46 | 11 | 269 | 11 | 13.95 | 38.43 | 53 |
| 14 | public health | 45 | 36 | 33 | 28 | 28 | 58 | 35 | 263 | 28 | 10.69 | 37.57 | 58 |
| 15 | clinical ethics | 46 | 35 | 17 | 29 | 25 | 64 | 43 | 259 | 17 | 15.59 | 37.00 | 64 |
| 16 | consent | 37 | 35 | 47 | 35 | 46 | 35 | 22 | 257 | 22 | 8.34 | 36.71 | 47 |
| 17 | privacy | 46 | 38 | 38 | 32 | 37 | 46 | 19 | 256 | 19 | 9.24 | 36.57 | 46 |
| 18 | advance directives | 34 | 34 | 37 | 30 | 37 | 40 | 42 | 254 | 30 | 4.03 | 36.29 | 42 |
| 19 | education | 57 | 35 | 29 | 25 | 41 | 48 | 16 | 251 | 16 | 14.03 | 35.86 | 57 |
| 20 | euthanasia | 41 | 29 | 22 | 25 | 38 | 49 | 36 | 240 | 22 | 9.52 | 34.29 | 49 |

6. Here it gets very nerdy

Notebook 3 – Interactive data exploration (Bioethics subset 2015 - 2021)

Select the entities to plot, click on 'add', then generate the plot.

Again, the input field is dynamic, so you can start typing and it will suggest existing entities.

You can see frequencies or normalized frequencies (click on the appropriate tab).

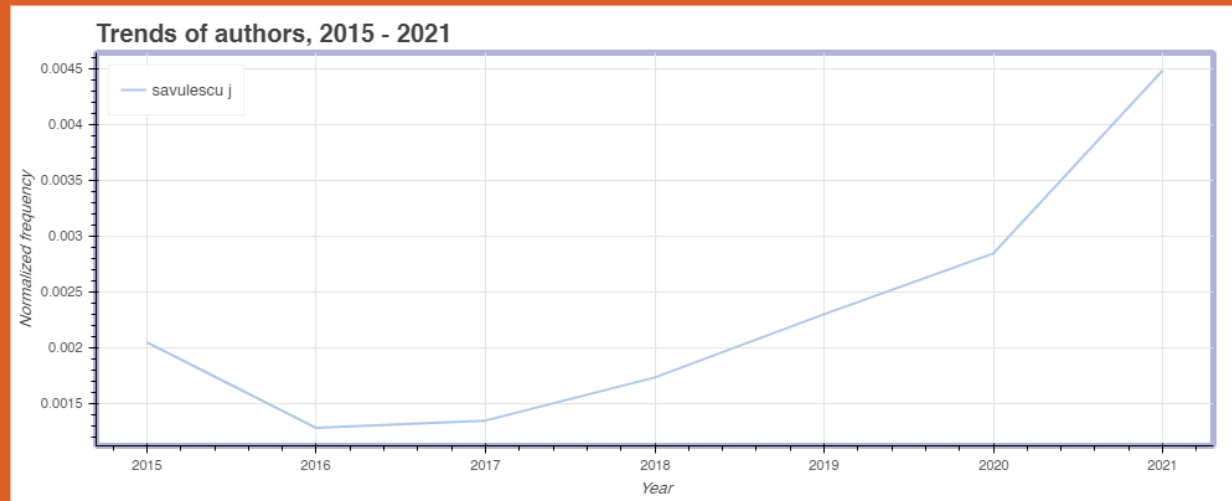
Hovering on a data point you'll see a hoverbox with further details.

Did I write more papers than my boss? Nope – but I wrote plenty of code!



For what is this useful?

- For query expansion – in a more refined way, differentiating between lemmas, keywords and MeSH;
- For plotting the trend of a concept in a field over time;
- For finding someone with extensive experience in some very specific field (you might want them at a conference, or as external supervisors, or as suggested reviewers, ...);
- For pre-mapping the topics you're going to find in a literature review;
- For picking the right journal (more on this topic later)
- *For bragging with your mom as soon as you finally publish more than your boss...*
- *... and for a bath of humility when you see that even at the very top of the pyramid your normalized impact on the field is still (quantitatively) insignificant.*



6. Open is cool

6. Open is cool

<https://www.uzh.ch/en/researchinnovation/openscience.html>

Open Science at UZH:

[LERU](#) member since 2006 (league of European research universities)

[DORA](#) signatory since 2014 (San Francisco Declaration on Research Assessment)

LERU [roadmap to open science 2018](#)

[Swissuniversities: Open access strategy 2017](#), Open research data strategy (end 2020)



We're open.
Are you?
[openscience.uzh.ch](https://www.uzh.ch/en/researchinnovation/openscience.html)

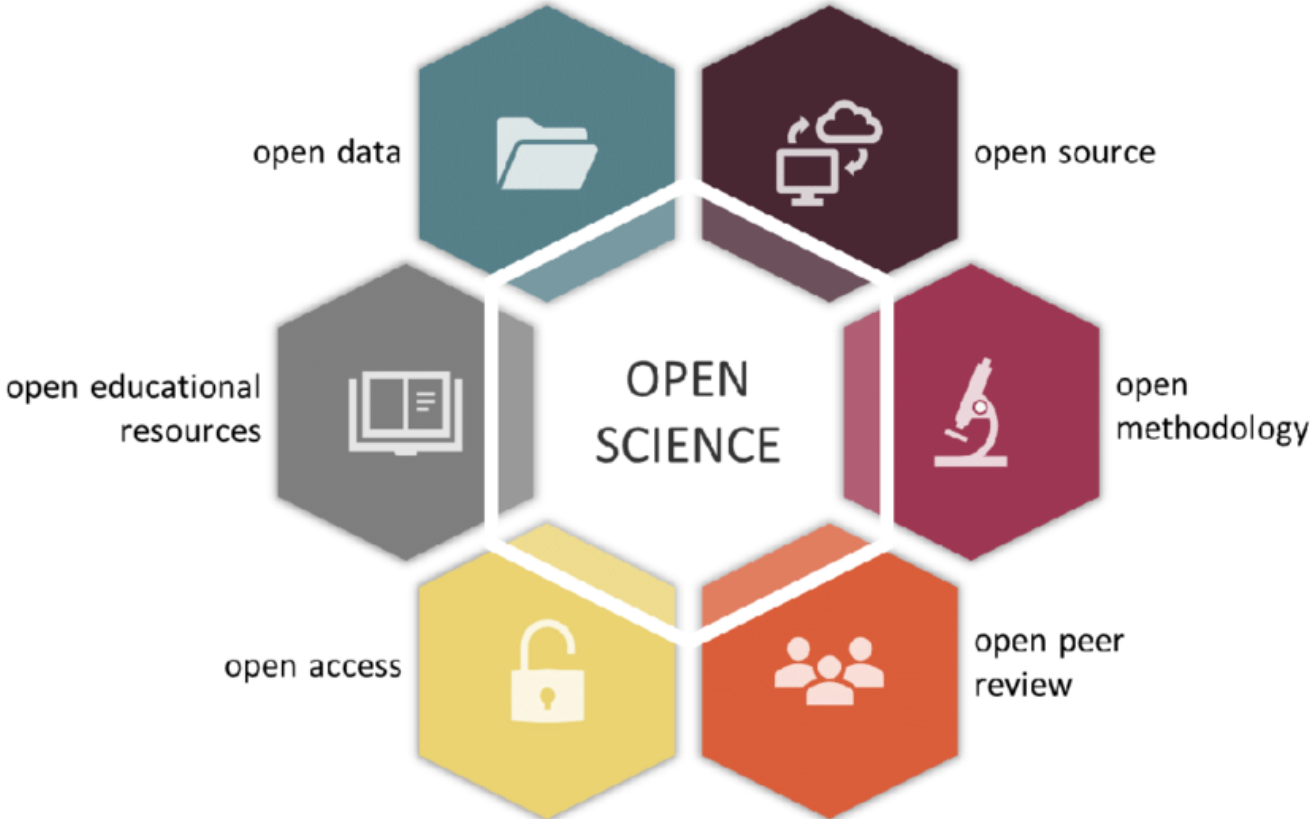
6. Open is cool

- 2011: data has to be shared (Funding Regulations)
- September 2015: the SNSF discussed the foundations of Open Research Data strategies during an international workshop
- 2016: Discussions at the Presiding Board and Administrative Offices
- October 2017: the policy enters into force in the project funding scheme.
- Data Management Plans (DMP) are now required in most of SNSF funding schemes
- Funding Regulations Article 47 on publication and accessibility of research results :
«data collected with the aid of an SNSF grant must be made available to other researchers and integrated into recognized scientific data pools»



**FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION**

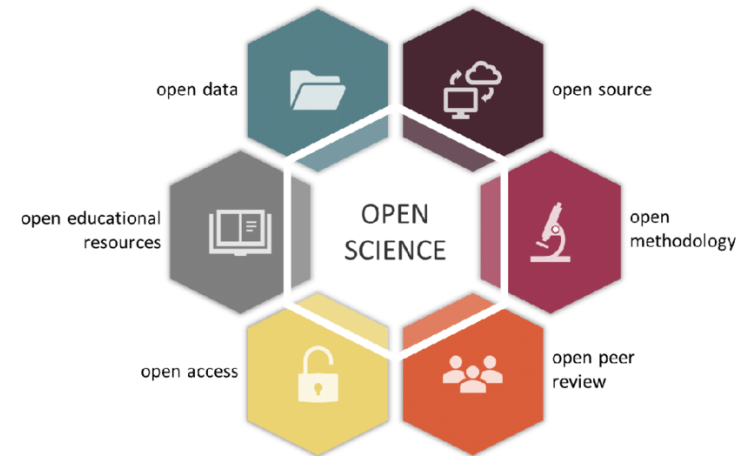
6. Open is cool



6. Open is cool

Search:

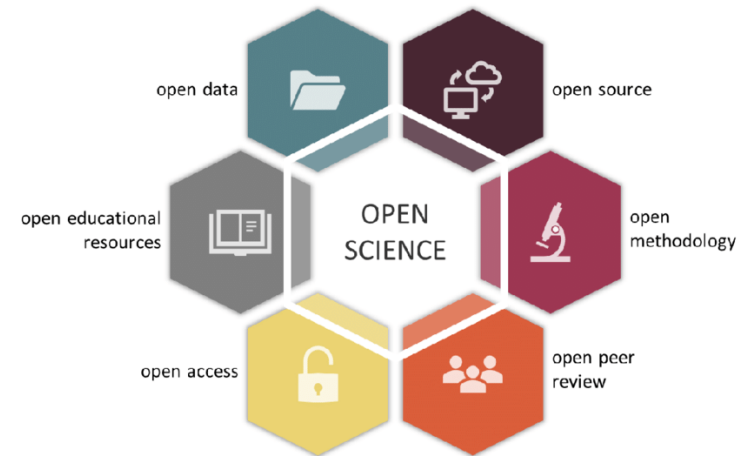
- Use pre-existing datasets (e.g. open data repositories like [Zenodo](#))
- Use shared reference libraries (e.g. [Zotero](#))
- Share grant proposals (e.g. at [RIO](#))
- Read and search OA journals ([DOAJ](#) or [Sherpa/Romeo](#))
- Involve [citizen science organizations](#) (if possible)
- Make use of wikimedia projects or contribute to them (e.g. wikipedia, wikidata, etc.)



6. Open is cool

Analysis:

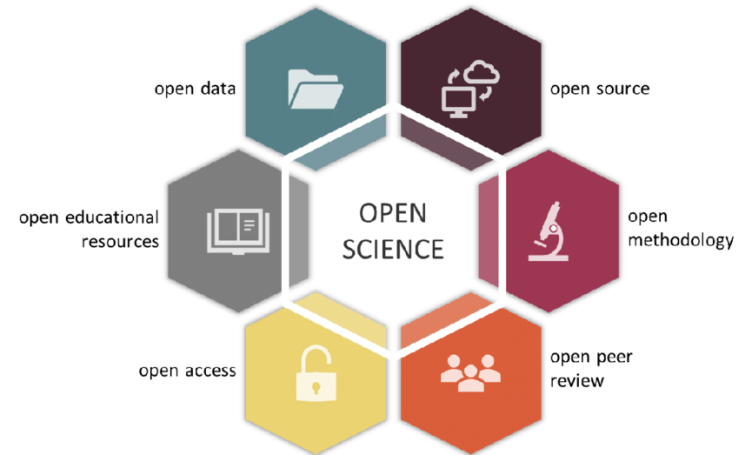
- Pre-register your research (e.g. Registered Reports, or [OSF](#))
- Share protocols and workflows (e.g. at [protocols.io](#))
- Share notebooks (e.g. [OpenNotebookScience](#))
- Share code, (e.g. via Github, as Jupyter notebooks)
- Share data (e.g. via Dryad, Zenodo, or Dataverse; see [re3data.org](#) for repositories)
- Make a data management plan
- Use open-source software and open formats



6. Open is cool

Writing:

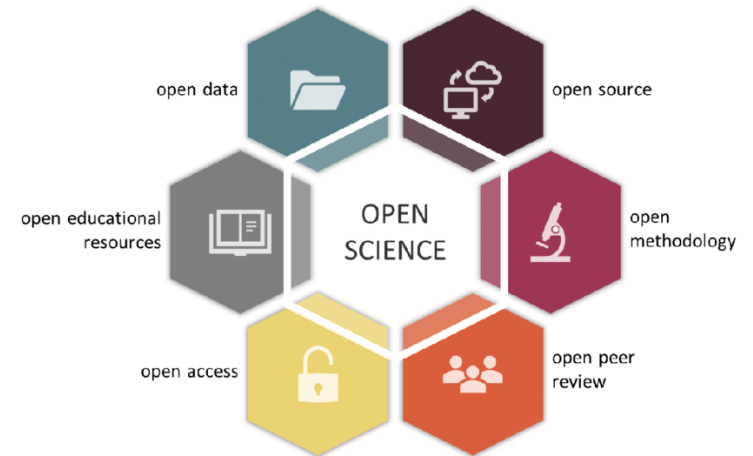
- Open XML or OpenDocument drafting
- Use actionable formats when mixing code and text, e.g. Jupyter or Markup / Latex
- Make use of tools such as wikidata, [open refine](#), etc.
- Include citations for software / datasets
- Discuss findings outside the institute before publishing



6. Open is cool

Dissemination:

- Advocate for open science
- Collaborate with researchers who practice open science
- Use social media or other platforms to talk about your work
- Use open science identifiers (DOIs and [ORCIDiDs](#)) for yourself and all your work
- Inform the wider public / community about your research (e.g. conferences)
- Involve Citizen Science organizations
- Publish your preprints (e.g. on Zenodo or OSF)
- Publish open access!



6. Open is cool

Benefits of Open Access



6. Open is cool



+ BLACK ROUTE
(legal in CH!)



The basics:

- Register an ORCID
- Use open data (when available) and release your data (Zenodo)
- Pre-register your research and share your protocols (OSF, Zenodo)
- Use open software (LibreOffice, Zotero, R, Python, ...)
- Archive your preprints (OSF, Zenodo)
- Publish on OA journals (DOAJ, Sherpa/Romeo)

7. Pick a journal

7. Pick a journal

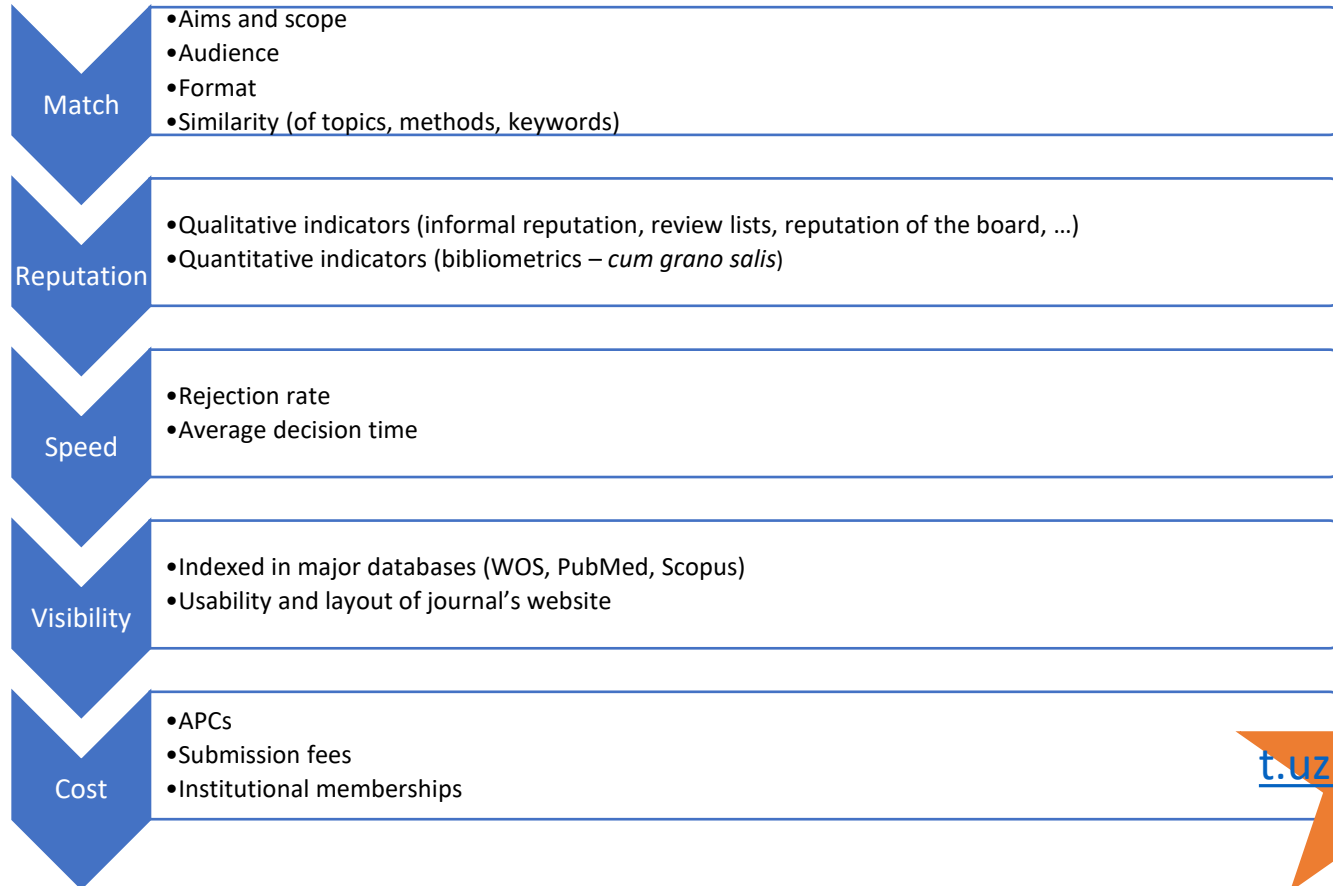
From DORA:

- citation distributions within journals are highly skewed;
- the properties of the Journal Impact Factor are field-specific: it is a composite of multiple, highly diverse article types, including primary research papers and reviews
- Journal Impact Factors can be manipulated (or “gamed”) by editorial policy;
- data used to calculate the Journal Impact Factors are neither transparent nor openly available to the public
- the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations;
- the need to assess research on its own merits rather than on the basis of the journal in which the research is published; and
- the need to capitalize on the opportunities provided by online publication (such as relaxing unnecessary limits on the number of words, figures, and references in articles, and exploring new indicators of significance and impact).





7. Pick a journal



t.uzh.ch/1eb

7. Pick a journal

Match:

<https://bioethics.georgetown.edu/bioethics-journals/>
<https://endnote.com/product-details/manuscript-matcher/>
<http://jane.biosemantics.org>
<https://journalfinder.elsevier.com>
<https://www.journalguide.com/>
<https://journalsuggester.springer.com>

TopicTracker

Reputation:

<https://beallist.net>
<https://clarivate.com/products/journal-citation-reports/>
<https://www.elsevier.com/solutions/scopus/how-scopus-works/metrics>
<https://www.journalindicators.com/indicators>
www.metrics-toolkit.org
<https://academic.microsoft.com/journals>
www.scimagojr.com/journalrank.php
<https://thinkchecksubmit.org>

Speed:

<https://scirev.org>
www.journalguide.com
<http://journalreviewer.org>

Visibility:

<https://doaj.org>
<https://v2.sherpa.ac.uk/romeo/search.html>

Costs:

<https://doaj.org>
<http://www.eigenfactor.org/openaccess/>

Thanks to Dr. Philipp Mayer for the input
<https://science-textflow.ch>

8. Discussion

SUMMARIZING:

- Amount of available literature
- Smart iterative search strategies
- Optimizing retrieval
- Content mining with MaxQDA
- TopicTracker
- Basics of open science
- Pick a journal (in a conscious way)

THANKS FOR YOUR TIME!

The owl of Minerva is watching you.



University of
Zurich^{UZH}

Institute of Biomedical Ethics
and History of Medicine