Information Literacy for your PhD
Successfully search, use and manage scientific publications

PhD beginners

2015-11-05

Thomas Henkel, Université de Fribourg
thomas.henkel@unifr.ch
Sylvie Vullioud, Scientific Information School (SIS)
sylvie.vullioud@scientific-info-school.ch

This document is largely based on PhD EPFL seminar support *"Scientific and Technical Information Literacy", Fall 2014 by originally published with CC-BY-NC- SA license and written by Alain Borel, Raphaël Grolimund, Julien Junod, Caroline Salamin and Sylvie Vullioud-Marcacci*

You have the right to reuse the content of this document *as long as this is not for commercial purposes and on condition that the documents derived from it shall also be under CC-BY-NC-SA license. In all cases, our names should be mentioned. For more detailed information about the license, go to the link indicated in the legal information below.*

Thomas Henkel and Sylvie Vullioud

Output for the participant

This seminar helps participants successfully search, manage and reuse the scientific information with which to begin their doctorate, or to update the publication list before they begin to write up their thesis.

General and specific search tools are addressed to enable the participants to better define the topic of their PhD thesis (has it already been covered?), to get to know their subject in depth (what is the state of the art in a particular field of research?), and to begin monitoring relevant scientific material (what are the latest publications on a topic?). The workshop introduces a reference management tool which permits the structured organization of the participant's documents and the easy citation of their sources.

We will also review current practices in bibliometrics and copyright law, with a view to participants' future publications

At the end of the seminar participants will be able to:

- find scientific publications effectively and efficiently
- use a reference management tool
- adapt their reuse of the sources of scientific information (texts, images, pictures, data) according to the different dissemination contexts of a document

Seminar program

9h00	9h15	**Welcome**	Welcome Seminar program presentation
9h15	09h45	**Initial assessment**	Icebreaker Expectations
09h45	10h00	**1. Scientific communication**	*Who are major academic publishers?*
10h00	11h00	**2. Search I**	*How to make the state of the art-on and find major publications related to a search question?*
11h	11h15		***Break***
11h15	12h	**3. Search II**	*Where do I find literature gems?* Group 1 Social Sciences & Humanities Group 2 Biomedical and exact sciences
12h	13h		***Lunch***
13h00	14h45	**4. Reference management**	*How do I organize my documentation?* Group 1 Social Sciences & Humanities: Citavi Group 2 Biomedical and exact sciences: Zotero
14h45	15h00	**5. Scientific watch**	*How do I start scientific watch?*
15h00	15h		***Break***
15h	15h45	**6. Citation and Copyright**	*What do I cite? And what should I be aware of?*
15h45	16h15	**7. Bibliometrics**	*What are Impact Factor (IF) and h-Index?*
16h15	16h45	**Final assessment**	Afternoon and day reformulations Check of expectations
16h45	17h00		Seminar evaluation

1. Scientific communication

What's up in the academic publication business?

09h45	10h00	15’

Section objectives

1	The participant is aware of major scientific **publishers**	
2	The participant is aware of **subscription** and **Open access** publications business models	
3	The participant is aware of what is **copyright transfer**	

1.1 Business of publishing Science Technical and Medical (STM) journals

In the 1960s and 1970s, commercial publishers began to selectively acquire "top-quality" journals which were previously published by nonprofit academic societies. Recently, merging also occurred

2004 Informa bought by Taylor & Francis

2005 Masson bought by Elsevier

2006 Blackwell (dummies collection) bought by Wiley

2009 CRC Press bought by Taylor & Francis

2014 Nature Publishing Group bought by Springer

So the scientific publishing market is dominated by the so-called ‘big four’ companies that comprise around 30% to 40% of the world’s total scholarly peer-reviewed journals, which number is about 40’000 journals titles according to Ulrichsweb, an online directory of scholarly journals:

Springer Science and Business Media has about 3’000 journals

Reed Elsevier has about 3’000 journals

Wiley & Sons has about 2’300 journals

Taylor & Francis has about 2’100 journals

There follows an enormous number of smaller publishers.

STM journal publishing is business and is estimated to be 9 to 12 billion $ annually**.** Business was mainly built so far on the copyright transfer to the publisher, resulting in need of subscription to access information. Now, the new publishing Gold-OA business model is rapidly increasing: Springer owns Biomed Central, one the biggest biomedical Gold-OA group of journals, that is a direct competitor to the independent PlosOne Gold-OA biomedical journals. Springer recently bought Nature Publishing Group, who just started to issue Nature Communication Gold-OA journal, direct competitor of Science Advances launched in 2014 by AAAS publisher. In Gold-OA model, author is charged 50 to 6000 $ to publish article, but any user can access freely the article. This Author Processing Charge (APC) per article could also be a good business for publishers, since pure electronic journals could publish an “infinite” number of articles, with special internet platform allowing some of automatic and faster reviewing of articles.

The income of scientific publishers comes from tax payers for subscription based Gold-Open Access journals (**Figure 1**). Both subscription and Gold journals may have an infinity of business models that can be classified in several categories: subscription-based journals, delayed-, Gold-, and Hybrid Open-Access journals (**Table 1**).

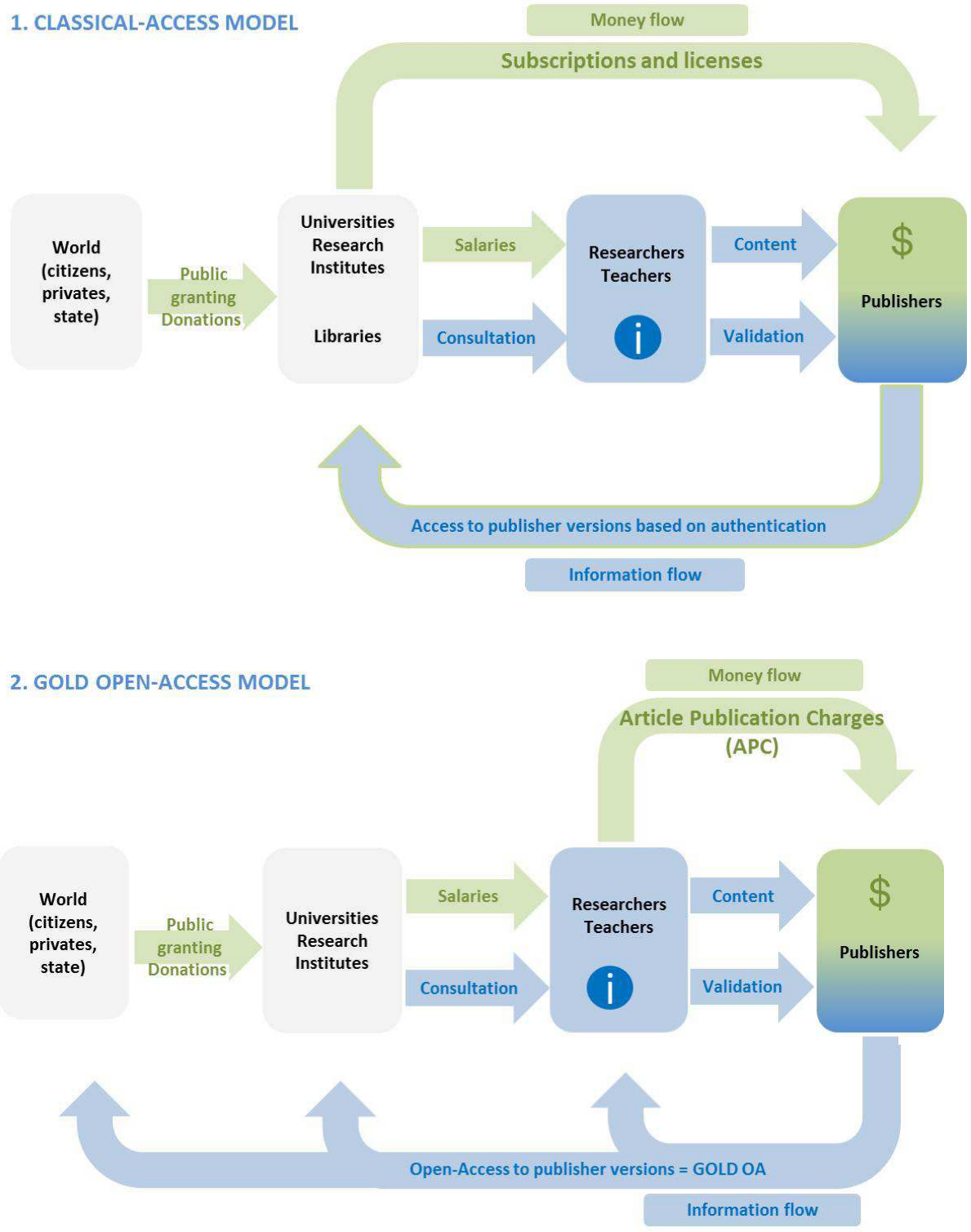


Figure 1 Subscriptions and Gold Open Access journals

A copyright transfer agreement is a part of commercial contract signed between authors and publishers, that have been historically facilitated the handling of copyright-based permissions in print-only publishing, and that is still at the base of business income of subscription based journals. Copyright transfer to publisher form author means that publisher only has the right to copy and diffuse scientific articles and books. In the age of electronic communication, the benefits of copyright transfer agreements have been questioned, and while they remain the norm, open licenses as used in open access publishing have been established as an alternative: one example is the form of the Creative Commons licenses: it allows anyone (including the publisher) to reproduce and distribute the work, with some possible restrictions. Creative Commons licenses are used by many gold open access journals, but not all.

Table 1

Journal category	**Business model**	
Subscription based	and/or	- Online **latest issues** and/or **back files** are paid by scholar libraries (**site licenses**, with IP or shibboleth identification) - Single user pays for one article (**Pay-per-View**) (average 40$ per article) - **Personal subscription** for online access journal - Subscription for online access journal only on dedicated local computer(s) - Subscription for online access journal on local computer(s) restricted to community only
Delayed-OA	and/or	- Online latest issues paid by scholar libraries (site licenses) with IP recognition or shibboleth identification After a period of time, back files become open access on publisher website - Single user pays for the latest article (**Pay-per-View**) (average 40$ per article) - **Personal subscription** to latest issues of online journal
Gold-OA	and	- The author (or laboratory and/or library and/or institutional research office and/or public research funder) pays to publish, allowing immediate access to the publisher version on the journal website (50 to 6000 $ per article) (- The author pay extra fee for CC-BY license) (- The author pay for possible self-archiving of publisher version manuscript)
Hybrid-OA	and/or	- The business model is still based on subscription journal paid by libraries but authors may choose **APC** (Author Publishing Charges) to allow immediate Open Access of their article. - Single user pays for one article (**Pay-per-View**) (average 40$ per article) - **Personal subscription** to the journal

1.2 Scientific search tools types

About 1.5 mio scientific peer-reviewed articles are published each year in subscription based, delayed, hybrid, and Gold journals. Search tools enable researcher to search scientific information, but because of so many business models, bibliographic data bases (BDB) or search engine can be very different in term of access and indexed type of full texts if any, and providers. In fact, if articles are not indexed by major search tools such as Google scholar, Web of Science, Scopus and subject specific search tools, they cannot be retrieved by internet users, therefore they have a low impact on research

2. Search I: good practices

How to find major publications related to a search question?

10h	11h	60’

Section objectives

1	The participant knows about differences and common points of **multidisciplinary search tools** (Google scholar, Web of Science, Scopus)	
2	The participant is able to **identify terms** describing a research question, and combines them with **Boolean operators**	
3	The participant knows to **fine tune** his search method if no results or too many results are retrieved	
4	The participant is able to (critically) **analyze** the content of different **scientific search tools**	

2.1 What is behind a search box? Keywords retrieval and interpretation / ranking of results *Before typing keywords*

The keywords of any search equation are retrieved by search engines which harvest web pages such as Google scholar, or bibliographic databases such as Nebis, Pubmed or ScienceDirect (**Figure 2**). Depending on the used Bibliographic Data Base (BDB), the search engine retrieves keywords from metadata, and/or abstracts, and/or chapter titles and/or full texts (**Tables 4, 5, 6**). Metadata is information about the document that is not part of the full text, i.e. article or book title, publication year, publisher, author, ISBN, ISSN, DOI, controlled vocabulary or author keywords. The Abstract is sometimes considered as metadata, sometimes not.

At the present time, keywords automatic interpretation occurs in the search box of almost all search tools in simple search mode: word variants are detected, such as singular and plural, adverb, noun and adjective, British and American spelling. This process is called *lemmatization* or *stemming*. It is performed automatically in Web of Science and Scopus for example. However, PubMed interprets a basic search query using tables of MeSH terms (controlled vocabulary within the bibliographic database), journal titles, author and investigator names, ending with complex search equation. This process is called *automatic mapping*. Sometimes automatic *lemmatization* and *mapping* lead to a better total recall of results, but with less precision. To keep control on the query, automatic interpretation can be disabled by using truncations in the search equation.

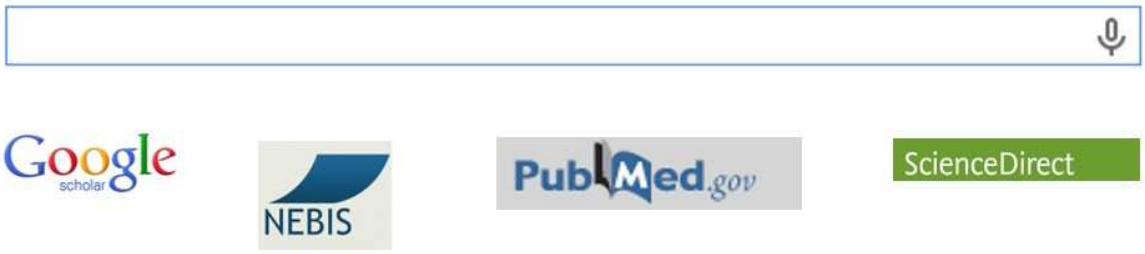


Figure 2 Search boxes look like all similar

The display of results by relevance relies on different calculations: in ScienceDirect and Engineering Village, it relies on static ranking, including location of term in title only, title and abstract, abstract and full text, frequency and proximity of terms, whereas Google scholar uses static ranking *and* dynamic ranking (popularity = number of links to a page) *and* citations counts

Most of the time, the exact algorithm of a search engine is not known because it is a commercial product. However, research has shown that Google scholar puts high weight especially on citation counts and words included in a document's title. As a consequence, the first search results are often highly cited articles. Google scholar seems to be more suitable for finding standard literature than gems or very recent literature. These results enable researchers to use further academic search engines and databases.

*Table 4 Tools and operating modes of some **multidisciplinary Bibliographic Databases** (BDB) and search engine*

Tool	**Type and use**	**Content**	**Keywords retrieval** ***Automatic interpretation***
Google scholar F	Multidisciplinary **citing** scholar search **engine**	Web addresses of peer-reviewed, conference **articles** and **patents** from **unknown resources,** including also **Predatory journal** articles**!**	**Full text, URL, metadata of webpage description, webpage title** *Unknown*
Web of Science	Multidisciplinary **citing** BDB for **fundamental** research	Selected «**best**» references of peer-reviewed and conference articles from **known resources**	**Metadata** (article title, abstract, keywords, keywords plus *None*
Scopus	Multidisciplinary **citing** BDB for **fundamental** research	Selected «**best**» references of peer-reviewed and conference articles from **known resources**	**Metadata** (article title, abstract, keywords) *Lemmatization*
Engineering Village	**Platform** dedicated to **engineering** possibly including Inspec, Compendex, Geobase and Georef BDB	**References** of peer-reviewed and conference articles from **known resources**	**Metadata** (article title, abstract, keywords) *None*
Internationale Bibliographie der geistes- und sozialwissenschaftlichen Zeitschriftenliteratur (IBZ)	**Citation BDB** in the humanities and social sciences	**References** of journal articles from **known resources**	**Metadata** (article title, keywords, German and English controlled vocabulary)
Dissertation Abstracts (Dissertation and Theses)	**Citation** BDB	**References** and **Abstracts** of **Doctoral Theses** from American, British, and other Universities	**Metadata** (article title, abstract, keywords, controlled vocabulary) *None*
Periodical Index Online Periodical Archive Online	**Citation** BDB with some access to full texts for social sciences and humanities	**References of articles** form **older journals (before 1996)** in the social sciences and humanities	**Metadata** (article title, keywords) *None*
Francis	**Citation** BDB in the humanities and social sciences especially with French speaking resources (but not exclusively)	**References** of journal articles from **known resources (by 2015!)**	**Metadata** (article title, keywords, French, English and Spanish controlled vocabulary) *None*
Humanities Index	**Citation** BDB in the humanities	**References** of English journal articles	**Metadata** (article title, keywords, controlled Vocabulary)
Publisher BDB			
CAIRN, IEEE Explore JSTOR ScienceDirect SpringerLink Taylor & Francis Wiley	Multidisciplinary **full text publisher BDB**	**References** of **one publisher** and corresponding **full texts**. Libraries can subscribe to entire or partial publisher reference BDB or entire reference BDB with partial access to full texts.	**Full text** and/or **metadata** (article title, abstract, keywords) *None*

*Table 5 Tools and operating modes of some **subject specific Bibliographic Databases** (BDB)*

Tool	**Type and use**	**Content**	**Keywords retrieval** ***Automatic interpretation***
Avery index, Iconda, International Bibliography of Arts	Architecture, Arts	**References** of peer-reviewed and conference **articles** from **known** specialized journals from different publishers. Some time, references of **books** and **book Chapters** are included (especially in the social sciences and humanities).	**Metadata** (article title, abstract, keywords, controlled Vocabulary with or without Thesaurus) *None*
Medline, Pubmed F, Embase	Biomedical		
Scifinder	Chemistry		
Transport	Civil engineering		
Geobase, Georef	Environment		
Business Source Premier, EconLit, WRDS	Economics & Management		
ERIC, FIS Bildung	Education		
Historical Abstracts, Brepolis Medieval Bibliographies	History		
MLA International Bibliography	(Modern) Languages and Literatures		
MathsciNet, ZentralBlatt	Maths		
Proquest Material Science	Materials		
Philosopher's Index	Philosophy		
Worldwide Political Sciences Abstracts	Political Sciences		
PsychINFO, PSYINDEX	Psychology		
Sociological Abstracts, International Bibliography of the Social Sciences	Social Sciences		
Urbadoc	Urbanism		

*Table 6 Tools and operating modes of some **Open Access Bibliographic Databases** (BDB) and search engines*

Tool	**Type and use**	**Content**	**Keywords retrieval** ***Automatic interpretation***
Open Institutional Repositories			
Archive ouverte F	UniGE	**Institutional-OA archives** of references **with** or **without** access to publisher, **pre and post referee full texts**	**Metadata** (article title, abstracts, keywords, **and not the full text!**) *None*
Infoscience F	EPFL		
Publication et Recherches F	UniNE		
ReroDOC F	UniFR		
Serval F	Unil		
Open Subject Repositories			
ArXiv F	Physics, Maths, Computing, Statistics	**References** of peer-reviewed and conference articles with or without access to **pre, post, publisher full texts**	**Metadata** (article title, abstract, keywords, **and not the full text!**) *Mapping for Pubmed Central tools / none*
US Pubmed Central F Europe Pubmed Central F Canada Pubmed Central F	Biomedical		
PhilPapers F	Philosophy		
Open Institutional Repositoires Search Engines			
Base-search.net F	Content of some Institutional and Subject **Open Access Repositories**	**References** of peer-reviewed and conference articles with or without access to **pre, post, publisher full texts**	**Metadata** (article title, abstracts, keywords)**,** full text for certain resources (Isidore) *None*
Recherche Isidore F	Content of Institutional and Subject **Open Access Repositories** for **Humanities** and **Social Sciences** especially with **French** resources (but not exclusively)		
Union library Catalogs			
RERO F	Swiss French Universities	**References** of **books and journals** (and **articles** for Nebis and RERO Explore). Use the Inter Library Loan to obtain the documents or photocopies	**Metadata** (author, title, keywords, controlled vocabulary) *None*
NEBIS F	Federal and HES institutions		
SWISSBIB F	Swiss academic libraries		

2.2 What is behind search box? Citing multidisciplinary search tools comparison *Before typing keywords*

The most popular current awareness tools for fundamental research are Web of Science (WoS) and Scopus, in all disciplines. They make you able to perform searches by cited peer-reviewed articles and conferences. These BDB are called *citing bibliographic databases* **(Table 7).** Scopus produced by Elsevier, is the competitor of WoS produced by Thomson Reuters. Scopus and WoS have the same scope for publishing the best journals, but Scopus is more focused on American journals whereas Elsevier is European. Google scholar is now of first importance. Total coverage of Google scholar is unknown: Elsevier journals were not included before 2007, and the most recent years of the American Chemical Society journals are apparently still missing. Google scholar does not publish a list of scientific journals crawled, and the frequency of its updates is unknown. It is therefore impossible to know how current or exhaustive searches are in Google scholar.

Table 7 Citing multidisciplinary search tools as a source of citation counts

	Journal coverage	**Price**	**Indexation criteria**	**Display of results**	**Alerts**	**Provider**
Web of Science	Selection of best journals based on the controversial Impact Factor (IF)	CHF 55'000/year	Impact Factor (IF) journals considered as "best" journals"	Publication date by default. Times cited, relevance, first author, source title	Citation Subject Journal Author	Thomson-Reuters
Scopus	Selection of best journals based on unknown criteria	CHF 45'000/year	Unknown selection of "best journal"	Publication date by default. Times cited, relevance, first author, source title	Citation Subject Author	Elsevier
Google scholar	Google does not produce any list of journals allowing crawlers to index their contents.	Free	Every journal publisher allowing indexation by crawlers	Relevance by default. Publication date	Citation Subject	Google

Web of Science, Scopus, and Google scholar measure the "impact" of articles by measuring the number of citations related to an article. These search tools are called citing search tools (warning: citation search tools mean that the content of the database is a collection of references).

Today, impact measurement of researcher relying on the function of citing tools (number of citations or h-index) is facing competitors: Plos One, CiteULike, Mendeley, Likes/Shares/Posts on Facebook, Research Gate, ORCID, PubmedCommons etc develop their own metrics of impact researcher called Altmetrics, using viewing and downloading of articles, and number of activities of post-reviewing comments, number of followers and following peers internet measurement, etc.

Exercise 1: Discover your institutional available search tools 10’

1. Find your institutional databases list.
2. Does your institution provide access to WoS and/or Scopus?

UNIL: start from http://www.bcu-lausanne.ch/

UniGe: start from: http://www.unige.ch/biblio/

UniNe: start from http://www2.unine.ch/bibliotheque

UniFr: start from http://www.fr.ch/bcuf

EPFL: start from http://library.epfl.ch

Comments

Exercise 2: discover citation counting functionality and connect it with the coverage content of search tools 10’

1. Why do you need your neighbors to do this exercise?
2. What is the number of citations for the following references?
3. What number of citation is false?

	Web of Science	Scopus	Google scholar
Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, *45*(1), 197-245.			
Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1), 83–119.			
Pattabiraman, V. R., & Bode, J. W. (2011). Rethinking amide bond synthesis. *Nature*, 480(7378), 471-479.			
Siegel, R., Ward, E., Brawley, O., & Jemal, A. (2011). Cancer statistics, 2011. The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA: A Cancer Journal for Clinicians*, 61(4), 212–236.			

Comments

2.3 Some Bibliographic Databases advantages over Google scholar

BDB citation counts seem to be less manipulated than in Google scholar

Google scholar does not take retraction articles into account if journals have not embedded a watermark in the PDF.

Why has the number of scientific retractions increased in the 21st century? Infractions have become more common and/or infractions are more quickly detected. Better understanding of the underlying causes for retractions can potentially inform efforts to change the culture of science and to stem a loss of trust in science among the lay public[1]

Furthermore, constant search technology changes within Google scholar make it impossible to obtain replicable search results. But it's just this replicability that is one of the fundamental principles of scientific research. In consequence Google scholar shouldn't be used for systematic reviews.[2]

[1] Source: Oransky Y. (12.07.2013) Why has the number of scientific retractions increased? Retraction Watch Blog. On http://retractionwatch.com/2013/07/11/why-has-the-number-of-scientific-retractions-increased-new-study-tries-to-answer [accessed 27.03.2015]

[2] Source: Anderson, P. F. (23.01.2013). What's Wrong With Google scholar for "Systematic" Reviews [Blog]. On https://etechlib.wordpress.com/2013/01/23/whats-wrong-with-google-scholar-for-systematic-reviews [accessed on 23.03.2015]

2.4 From scientific question to search equation

A. Formulate a question

I want to know what strains of mice are used as models for Huntington studies, and especially if inducible transgenic mouse are existing?

B. Classify keywords from the most general to the most specific with a tree or grid[3]

Neurodegenerative disease,-s

Dementia

Huntington

Huntingtin

Animal experimentation, Laboratory animal, Animal model

Vertebrate model

Small animal model

Rat model

Mouse / mice model

Transgenic mouse

Inducible/ conditional / Tet-on / Tet-off/ Tet mouse

Topic	Huntington's disease modelling with inducible mouse		
Key Concepts	inducible transgenic mouse	Huntington	model
Synonyms	Conditional, tet, tet-on, tet-off mouse	Huntington's disease	Modelling, experimentation, laboratory
Broader Terms	Small animal models, vertebrate	Neurodegenerative disease	
Narrower Terms		Dementia	
Related Terms		Huntingtin	

C. Formulate some corresponding search equations

animal? AND (model? OR laborator*) AND (neurodegenerative OR dementia OR Huntington)

(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington

Miniguide for best precision and recall of results

AND Boolean operator for combination of concepts

OR Boolean operator for expressing variants, synonyms and close concepts

(…) brackets for nesting

"..." quotation marks wildcard for the search of expression

***** truncation wildcard for variants of a term roots (singular, plural, adjective, and adverb)

[3] Source: http://lotse.uni-muenster.de

2.5 Information research is an iterative process

What strains of mice are used as models for Huntington studies, and more specifically are there inducible transgenic mice existing?

Step	Keywords/equations	Tool	Number of results	Relevance of results	Comments
1	(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington	NEBIS	2	Bad	Keywords are too narrow. No book of 500 pages is written on inducible mice for Huntington study. *Use of more broad keywords to find a book about animal model of Huntington disease.*
2	animal? AND (model? OR laborator*) AND (neurodegenerative OR dementia OR Huntington)	NEBIS	110	Good	Use filter “books”! When looking for books, less narrow keywords give more results
3	(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington	WOS	18	Good	Few results. Further investigation is needed to increase number of results *- Work needed on variant spelling of tet-on, tet on, teton etc.* *- Inducible mice not very often used for Huntington study?* *- Need to add mouse strains specific terms.*
4	(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington	Scopus	22	Good	Same comments as n°3
5	animal? AND (model? OR laborator*) AND (neurodegenerative OR dementia OR Huntington)	Pubmed	11'072	Bad	Keywords are too broed to be relevant for articles retrieval
6	(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington	Pubmed	15	Good	Automatic mapping disabled, due to use of “...” . Need to increase number of results by using the mesh table
7	(Tet-on[tiab] OR tet-off[tiab] OR tet[tiab] OR inducible[tiab]) AND (mouse[tiab] OR mice[tiab] OR mice[mh]) AND (huntington[tiab] OR Huntington disease[mh])	Pubmed	46	Good	Same comments as n°3
8	(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice") AND Huntington	Science Direct	41'191	Bad	Search by default in full text gives much more results than before. Need to use strain names to reduce noise
9	(Tet-on[tiab] OR tet-off[tiab] OR tet[tiab] OR inducible[tiab]) AND (mouse[tiab] OR mice[tiab] OR mice[mh]) AND (huntington[tiab] OR Huntington disease[mh])	Pubmed Alert	1-2 / months	Good	Same comments as n°3
10	(Tet-on OR tet-off OR Tet OR inducible mouse OR inducible mice) AND Huntington	WOS Alert	1-2/ months	Good	Same comments as n°3

Exercise 3: turn your scientific question into a search equation and test it in a search tool to obtain best recall and precision

A. Formulate a question

B. Classify keywords from the most general to the most specific with a tree or grid

C. Formulate corresponding search equations

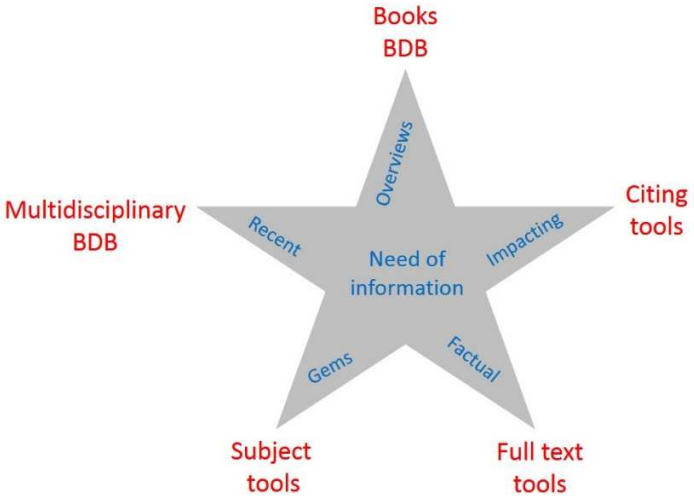
Mini guide for improvement of search results

Tool selection

- **Book BDB** to have a good **overview** on subject knowledge
- **Full text BDB** to find information located in the full text such as material, methods, or factual info)
- **Citing tools** WoS Scopus and Google scholar to identify most impacting journal articles
- **Specialized** licensed BDB or **Subject** Open Repositories to identify most "**niche**" and **gems** articles

Display of results

- By **date** to find the most recent articles
- By **relevance** to find the most relevant articles
- By **times** cited to find the most popular/important articles



Too many irrelevant results->
decrease the number of results

- Use **AND**
- Add or use **specific** terms
- (by adding too many AND in a keywords combination, you may end up with no results, if the search engine is based on a pure boolean retrieval system)
- Use **filters**
- Restrict to **metadata** search

Too few relevant results-> increase the number of results

- Use **OR**
- Add or use **broader** terms
- Use **full text** search tools
- (The subject is very recent and few things are published yet)

Exercise 4: test 2 equations for WoS, Scopus and Google scholar to find most cited articles and reviews 15’

Tested question =……..

Equation 1	……		
	Action	**Number of results**	**Relevant references example**
WOS	Display result by highest cited articles		
	Add the Review filter		
Scopus	Display result by highest cited articles		
	Add the Review filter		
Google scholar	Display result by highest cited articles		
	Add the Review filter		

Equation 2	……		
	Action	**Number of results**	**Relevant references example**
WOS	Display result by decreasing order of cited articles		
	Add the Review filter		
Scopus	Display result by decreasing order of cited articles		
	Add the Review filter		
Google scholar	Display result by decreasing order of cited articles		
	Add the Review filter		

Exercise 5: Discussion trainers and trainees: compare biomedical, exact and environmental sciences with social sciences practices

Do you think that the use of keywords and equations combined with WOS, Scopus and Google scholar is the best strategy for the finding major publications related to a search question in social sciences?

Comments

3. Search II: good practices

Where do I find literature gems?

11h15	12h	45’

Documentation apart this main course note support

Section objectives

Group 1 Social Sciences & Humanities
Group 2 Biomedical and exact sciences

1	The participant selects and uses relevant specialized search tools available at his institution for his area of research (**different platforms**)	
2	The participant is able to use **controlled vocabulary** of a subject database relevant for his domain	

Exercise 6: Discover your institutional available search tools 10’

1. Find your institutional databases list.
2. Which BDB are provided by your institution that may be useful for your PhD?

UNIL: start from http://www.bcu-lausanne.ch/

UniGe: start from: http://www.unige.ch/biblio/

UniNe: start from http://www2.unine.ch/bibliotheque

UniFr: start from http://www.fr.ch/bcuf

EPFL: start from http://library.epfl.ch

Comments

Exercise 7: Use the controlled vocabulary of the selected search tool for your topic search 20’

1. Which subject terms of the controlled vocabulary correspond to your keywords in your tree or grid (cf. exercice 3B)?
2. Use the thesaurus (if available) or the different subject indexes/fields provided by the database.
3. Analyze a single search result and identify other fields and terms serving to define your topic.
4. Use the identified terms and fields for your search.
5. Evaluate your search result and refine your search by:
   - using filters
   - identifying better subject terms
   - combining search results in the search history

Comments

4. Information management

How do I organize my documentation?

13h	14h45	1h45

Documentation apart this main course note support

Section objectives

1	Group 1 Social Sciences & Humanities The participant saves, cites, and organizes references with **Citavi**	
2	Group 2 Biomedical and exact sciences The participant saves, cites, organizes and share references with **Zotero**	
3	The participant is able to **report** his information research strategy	

4.1 Reference manager generalities

A reference manager does

- add/import references to your library very quick
- cite references while your write (almost) painless
- (share the references with your colleagues, advisor, lab, working group ...)
- (store the pdf files)

If you use a home-made database, you won't be able to share and you won't have any help to cite.

But a reference manager does not

- make sure the reference is complete and correct
- read the papers
- choose the relevant ones

Many reference managers are available. The **table 8** below gives an overview of four of them.

Table 8 Overview of Mendeley, Papers, Zotero and Citavi reference manager

Mendeley	**Papers**	**Zotero**	**Citavi**
free	59€ / $79 *40% discount for students*	free	149 CHF[4]
proprietary Elsevier	proprietary Springer	free Open source	proprietary Swiss Academic Software
Windows Mac OS X[5] GNU/Linux	Windows Mac OS X	Windows Mac OS X GNU/Linux	Windows
integrated with MS Word OOo / LO[6] LaTeX[7]	*integrated with* MS Word OOo / LO[9] LaTeX[10]	*integrated with* MS Word OOo / LO[9] LaTeX[10]	*Integrated with* *MS Word* *OOo / LO*[9] *LaTeX*[8]
7,100+ citation styles (CSL)	7,100+ citation styles[9] (CSL)	7,100+ citation styles[10] (CSL)	3000+ citation styles[11] (integrated style editor)

4 Free version for projects with up to 100 references, members of the University of Fribourg: CHF 10 (http://www.unifr.ch/micromus/fr/software/citavi)

5 "On Mac, the OpenOffice plugin was dropped in favor of support for LibreOffice" (source: http://www.mendeley.com/release-notes/v1_5/, accessed 24.02.15).

6 OOo = OpenOffice.org / LO = LibreOffice

7 .bib file creation needed (through export of references)

8 Citavi works with several TeX editors: https://service.citavi.com/KB/a201/32071-does-citavi-work-with-latex-documents.aspx (accessed 12.03.2014)

9 Source: http://www.papersapp.com/ (accessed 21.11.22014)

10 Source: http://zotero.org/styles (accessed 21.11. 2014)

11 Source: http://citavi.com/en/features.html (accessed: 12.03.2015), user request possible (http://citavi.com/sub/machform/view.php?id=47)

4.2 Search reporting

A reference manager is not only useful to cite while you write articles, reports, and thesis, but it can be also be used for keeping trace of a search methodology. This reporting may be useful to explain your thesis supervisor how you performed the state of art on a key question or how you kept up-to-date before your PhD exam. Moreover, search reporting may be useful for justification of search strategy for systematic reviews. Finally, in biomedicine, it may be also useful to justify the absence of animal experimentation alternatives when submitting an authorization form to ethical committees.

Example: reported equations and tools for search reporting

Keywords/equations	**Tools / Alerts**	**Number of results/ / relevance**	**Selected useful references**
animal? AND (model? OR laborator*) AND (neurodegenerative OR dementia OR Huntington)	NEBIS	1245 Acceptable	[1]-[3]
(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice" OR "conditional mouse" OR "conditional mice") AND Huntington	WOS	28 Publication date display / Good	[4]-[5]
	WOS	28 *Times cited* display / Good	[6], cited 606x!
(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice" OR "conditional mouse" OR "conditional mice") AND Huntington	Scopus	30	[7],[8]
(Tet-on[tiab] OR tet-off[tiab] OR Tet[tiab] OR "inducible mouse"[tiab] OR "inducible mice"[tiab] OR "conditional mouse"[tiab] OR "conditional mice"[tiab]) AND (huntington[tiab] OR Huntington disease[mh])	Pubmed	19	[9]
(Tet-on[tiab] OR tet-off[tiab] OR Tet[tiab] OR "inducible mouse"[tiab] OR "inducible mice"[tiab] OR "conditional mouse"[tiab] OR "conditional mice"[tiab]) AND (huntington[tiab] OR Huntington disease[mh])	Pubmed Alert	1-2 / months	
(Tet-on OR tet-off OR Tet OR "inducible mouse" OR "inducible mice" OR "conditional mouse" OR "conditional mice") AND Huntington	WOS Alert	1-2/ months	
[6]	WOS citation alert	1/week	Once per month is enough
To be determined	Journal alert		

Corresponding bibliography for mice models in Huntington studies

[1] Jesús Avila, *Animal models for neurodegenerative disease*. Cambridge: Royal Society of Chemistry, 2011.

[2] P. P. de Deyn, *Animal models of dementia*. New York: Humana Press, 2011.

[3] Siegfried Hoyer, Workshop « Cell and Animal Models in Aging and Dementia Research » (1993, et Heidelberg), *Cell and animal models in aging and dementia research*. Wien etc: Springer, 1994.

[4] Z. Ortega, M. Diaz-Hernandez, C. J. Maynard, F. Hernandez, N. P. Dantuma, et J. J. Lucas, « Acute Polyglutamine Expression in Inducible Mouse Model Unravels Ubiquitin/Proteasome System Impairment and Permanent Recovery Attributable to Aggregate Formation », *J. Neurosci.*, vol. 30, n° 10, p. 3675-3688, mars 2010.

[5] H. B. Kordasiewicz, L. M. Stanek, E. V. Wancewicz, C. Mazur, M. M. McAlonis, K. A. Pytel, J. W. Artates, A. Weiss, S. H. Cheng, L. S. Shihabuddin, G. Hung, C. F. Bennett, et D. W. Cleveland, « Sustained Therapeutic Reversal of Huntington's Disease by Transient Repression of Huntingtin Synthesis », *Neuron*, vol. 74, n° 6, p. 1031-1044, juin 2012.

[6] A. Yamamoto, J. J. Lucas, et R. Hen, « Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease », *Cell*, vol. 101, n° 1, p. 57-66, mars 2000.

[7] X. Gu, V. M. André, C. Cepeda, S.-H. Li, X.-J. Li, M. S. Levine, et X. William Yang, « Pathological cell-cell interactions are necessary for striatal pathogenesis in a conditional mouse model of Huntington's disease », *Mol. Neurodegener.*, vol. 2, n° 1, 2007.

[8] H. J. Han, C. C. Allen, C. M. Buchovecky, M. J. Yetman, H. A. Born, M. A. Marin, S. P. Rodgers, B. J. Song, H.-C. Lu, M. J. Justice, F. J. Probst, et J. L. Jankowsky, « Strain background influences neurotoxicity and behavioral abnormalities in mice expressing the tetracycline transactivator », *J. Neurosci.*, vol. 32, n° 31, p. 10574-10586, 2012.

[9] A. Saavedra, A. Giralt, L. Rué, X. Xifró, J. Xu, Z. Ortega, J. J. Lucas, P. J. Lombroso, J. Alberch, et E. Pérez-Navarro, « Striatal-enriched protein tyrosine phosphatase expression and activity in Huntington's disease: a STEP in the resistance to excitotoxicity », *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 31, n° 22, p. 8150-8162, juin 2011.

Exercise 8: search, save, cite at the same time for search reporting or the writing of a text

Comment

5. Scientific watch

How do I stay up-to-date?

14h45	15h00	15’

Section objectives

1	The participant stays up-to-date on a specific question effortlessly thanks to email **alerts** or RSS feeds (search, citation, author, journal, and comment alerts)	

In an ever evolving research field, you need to know what's new. Staying up-to-date is time-consuming. Among others, there are two ways to make it efficient. Both are made to pull information from its source to you, so that you don't have to visit all the interesting websites on a regular basis.

It's a 3 steps process

First, find a relevant query/journal/author

Secondly, subscribe to a feed or create an email alert.

And then... you can go to the beach!

Email alerts

Many scientific databases or editor platforms offer an email alert feature. Most of the time, you need to create an account, with the exception of Google scholar (**Figure 6**). Then, you can save queries or/and create alerts on a specific journal, author, subject (**Figure 7**)

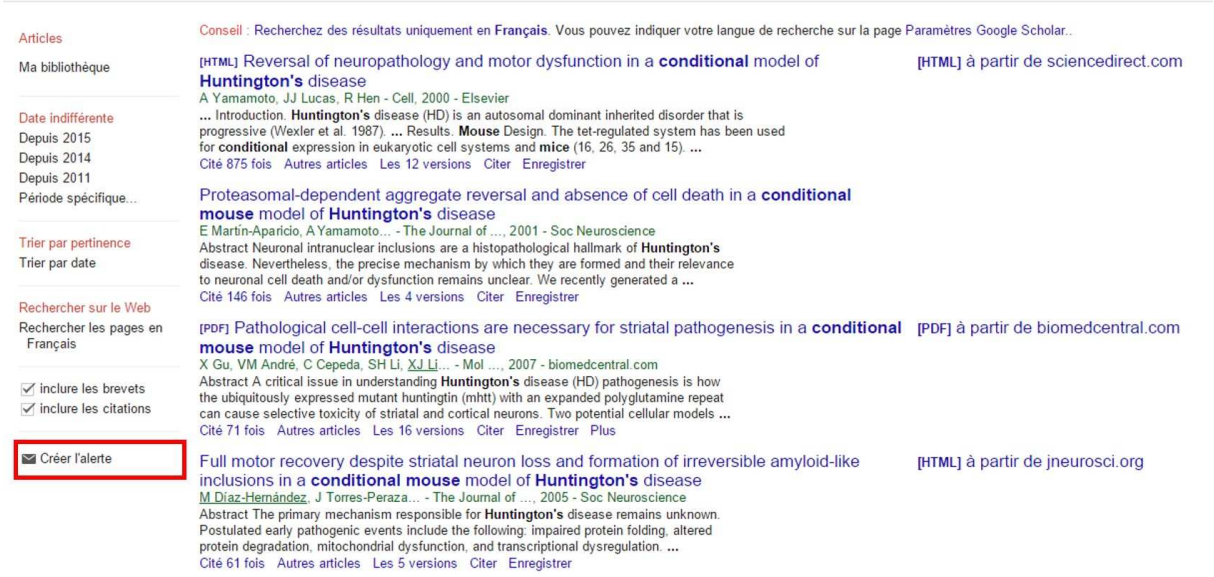


Figure 6 Google scholar subject alert without creation of an account

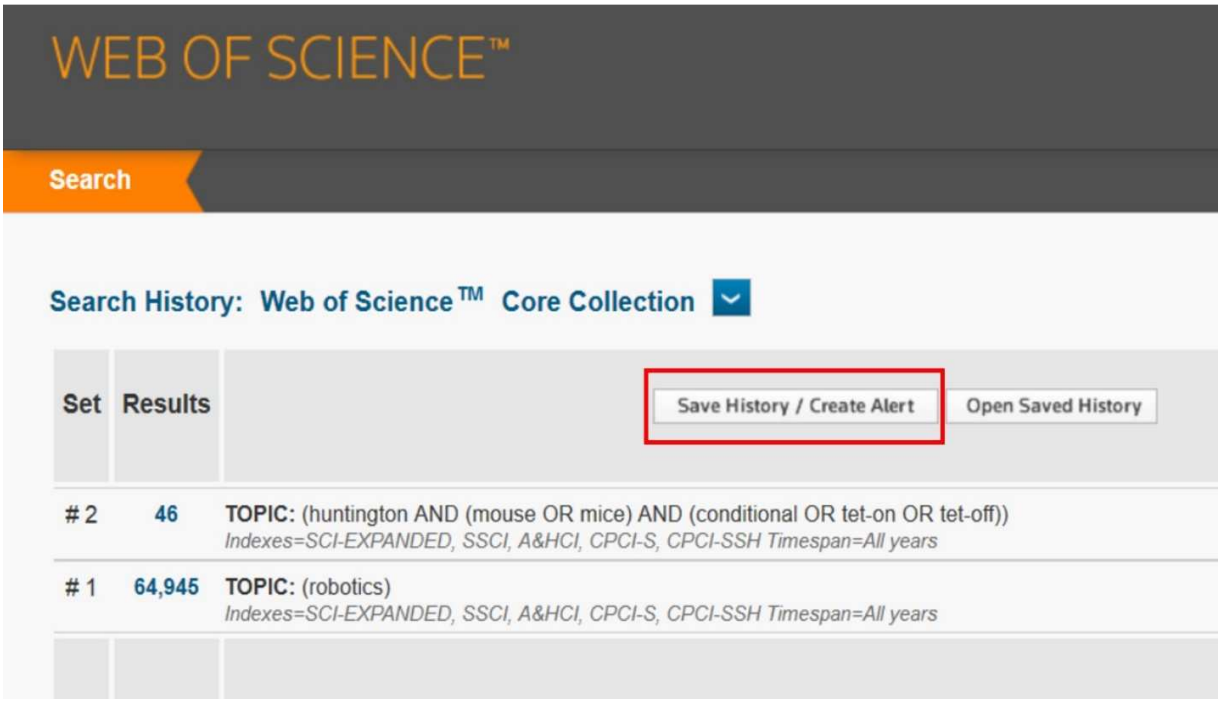


Figure 7 WOS subject and citation email alert

RSS feeds

RSS feeds bring you the same benefits as email alerts, but you don't need to create an account to subscribe to an RSS feed. A simple click on the RSS icon enables you to add the feed's URL to your reader. You can find both email alerts and RSS feeds on the same platform (**Figures 8, 9, 10, 11**)

To read an RSS feed, the best tools are aggregators. They offer advanced management features to store, tag, order, sort and search in a collection of feeds. But an email client or a web browser can read RSS feeds as well. They just lack advanced management tools. If you don't need them, you don't need a specific software to manage your feeds.

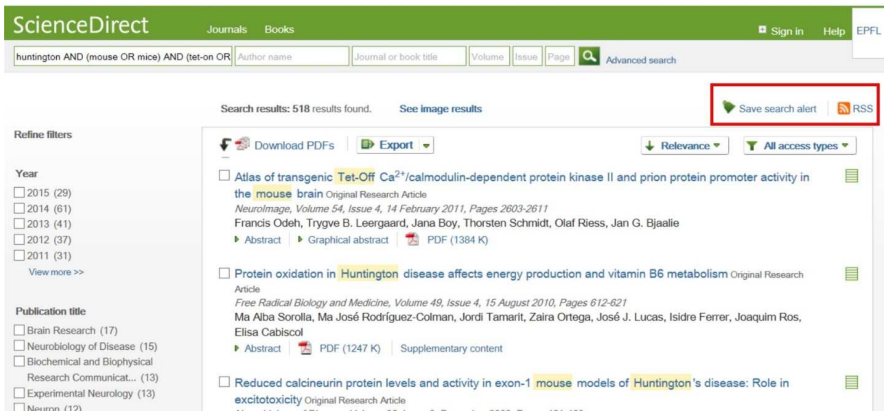


Figure 8 RSS feed and email alert from a query on Science Direct (Subject alert)

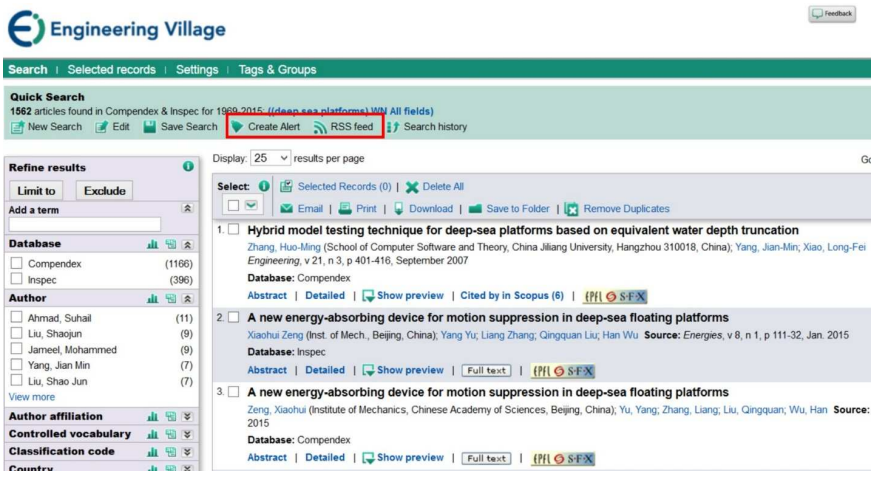


Figure 9 RSS feed and email alert from a query on Engineering Village (Subject alert)

Figure 10 RSS and email alert from a journal on IOP platform (Journal alert)

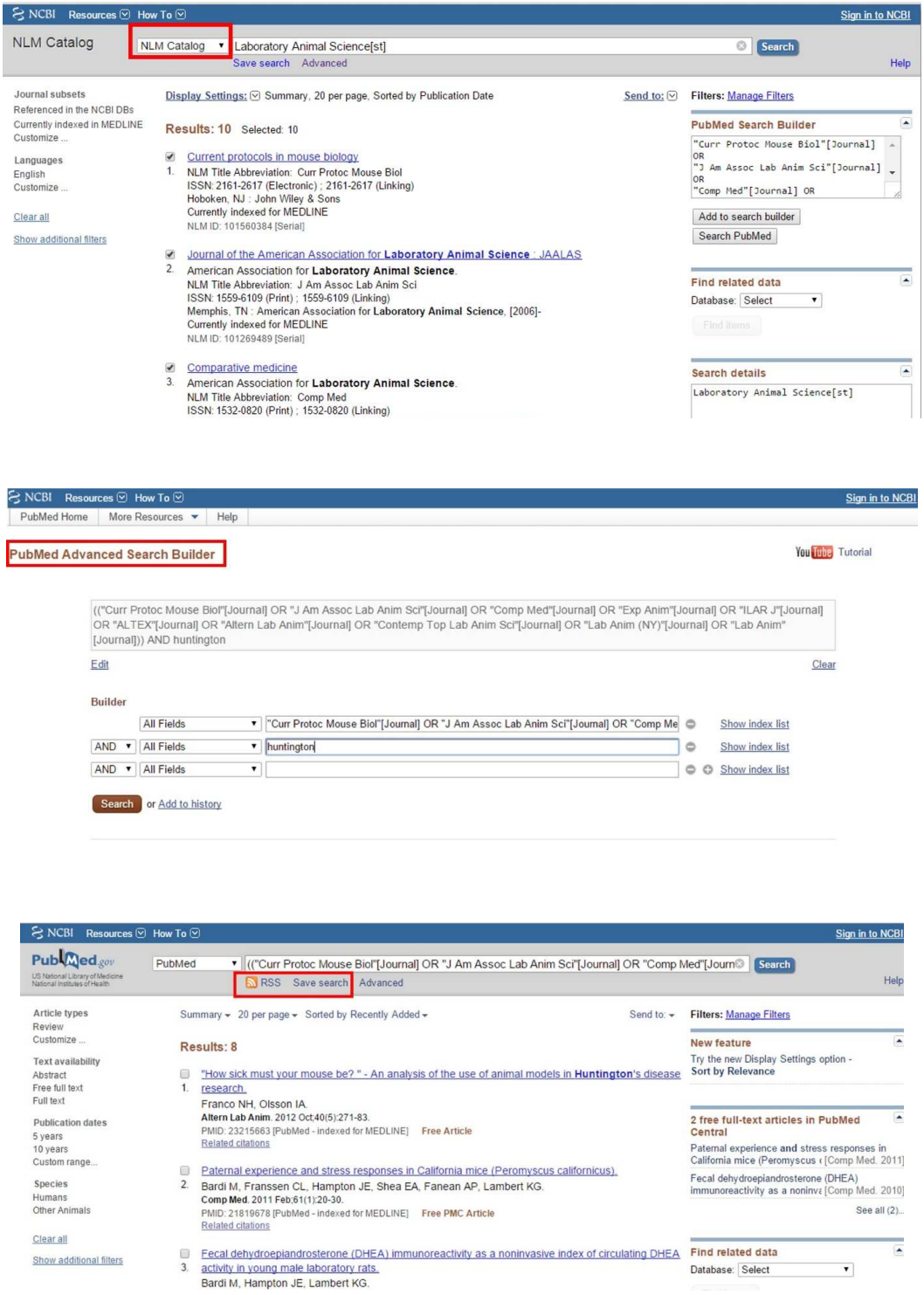


Figure 11 Main step for subject alert on journal selection in Pubmed

Miniguide of some strategies for scientific watch

Personal Network

- Supervisor
- Conferences
- Colleagues
- Lab colleagues bibliographies
- Group meetings

Push

- Citation alerts
- Laboratory journal-club
- Subject, journal, author, comment alerts
- Newsletters subscritpions

Pull

- Regular search in different databases and search engines
- Regular search on the most prestigious conferences websites

Social media

- Twitter
- LinkedIn
- ResearchGate
- ORCID
- Specialized blogs

And many others

Exercise 9: set up alerts

Comments

6. Citation and copyright

What do I cite? And what should I be aware of?

15h	15h45	45’

Section objectives

1	The participant knows **why to cite** and how to comply with copyright	
2	The PhD candidate is aware of what kind of **copyright compliance questions** should be raised for the writing of peer-reviewed articles or thesis	

6.1 Citation

Why do I cite stuff?

Miniguide: the 7 commandments of citation

1. I cite to acknowledge someone else's work
2. I cite to put my work in perspective
3. I cite to re-use existing work in my own
4. I cite to help the reader to discover a new source of information or to check it
5. **I only cite references that I read and understand**
6. I insert in reference as much as information publisher style guidelines allows
7. I should avoid citing references to please the boss/reviewer/publisher

What has to be cited?

- Any idea that is not yours should be properly cited
- Data collected and/or analyzed by others must be properly cited
- Figures, intact or derivatives, under Creative Common license must be properly cited
- Figures under Copyright should have the labelling of reproduction permission

What should not be cited?

- Trivial facts, and common knowledge should not be cited
- Avoid citing unpublished results or private communications (even if the original authors gave their blessing) unless absolutely necessary
- Use the references your supervisor gave you BUT MAKE SURE YOU'VE READ THEM
- Be ready to add suggested references after the first round of peer review, but find the appropriate spot for them
- Journal publishers have sometimes requested more citations to their own journal or other titles from the same publisher. Resist these attempts.

What kind of references to cite?

- Use background references, especially in the introduction section to make the paper more trustworthy
- Use context references to demonstrate the originality of the work
- Use theoretical and methodological bases references that were used the work. Avoid "it is well-known that"

Best practice of citing

- Use text with proper quotes, but not too much
- Direct quotes must be noted explicitly

  Example: According to Smith et al., « … [the results of this method] are not reliable ».
- Use a neutral tone to introduce a citation, regardless of your own feelings about it

- Citations must be attributed to the proper author
- Cite accessible references to obscure ones when they are more or less equivalent:

  Example: Case of translated article

  Choose

  N. Arutyunyan, A. Zaitsev, and N. Shaposhnikov, “Analyzing the phase composition of Si-B and Si-B-Ti alloys based on silicon,” Russian Journal of Physical Chemistry A, Focus on Chemistry, 86, 3, 339–341, 2012.

  Instead of

  Н. А. Арутюнян, А. И. Зайцев и Н. Г. Шапошников “АНАЛИЗ ФАЗОВОГО СОСТАВА СПЛАВОВ Si–B И Si–B–Ti НА ОСНОВЕ КРЕМНИЯ », Журнал физической химии, 86, 3, 405-408, 2012

- Use indirect citations with proper care; the original literature is a better choice in general

  Example: Case of indirect citation lead to scientific error

  Berger[12] raised concern about scientific validity of recent studies citing for decades a believed American reference[13] on the cold on human from 1946. Many physiologists cited this reference without knowing that it was not a primary source of information. The Nazis made experiments on prisoners without ethical considerations and with ideology causing scientific bias. The American reference is a secondary reference of the original work, but none of physiologist took time to access the real primary source of information.

Reference writing tips

General rule for reference writing is to insert as much information as your reference manager allows. Your citation management tool will do most of the work for you. Then insert as much as information publisher style guidelines allows. Be especially careful when you cite “exotic” document types (patents, conference papers). Rely on norm ISO 690 for the writing of references

Example: Peer-reviewed article formats required by different journal publisher

McCaffery, A. J. In Molecular Electronic Structures of Transition Metal Complexes II; D.M.P. Mingos; P. Day ; J.P. Dahl, Eds.; Springer: Berlin, 2012; pp. 121–148
=
McCaffery, A. J. From Ligand Field Theory to molecular collision dynamics: A common thread of angular momentum. Molecular Electronic Structures of Transition Metal Complexes II 121–148 (2012). at http://dx.doi.org/10.1007/10.1007/430_2011_51
=
McCaffery, A.J., Struct. Bond. 2012, 143, 121
=
MCCAFFERY, A.J., 2012. From Ligand Field Theory to molecular collision dynamics: A common thread of angular momentum. In: D.M.P. MINGOS, P. DAY and J.D. DAHL (eds.), Molecular Electronic Structures of Transition Metal Complexes II [online]. Berlin: Springer. Structure and Bonding, 143. pp. 121–148. [Accessed 11 April 2012]. ISBN 978-3-642-27378-0.

Example: proceedings article and patent reference

Smith J. What I did during my summer vacation. In Aebischer P, editor. EPFL Yearbook 2011. EPFL Homecoming Celebration; 2011 Sep 19; Lausanne. Berlin: Springer; 2012. p. 256-512

Smith Alain Robert A, Pamplemousse Editors. Automatic peer-reviewing. European patent EP20124815162342. 2012 Apr 1

[12] Source: R. L. Berger, “Ethics in Scientific Communication: Study of a Problem Case.,” J. Med. Ethics, vol. 20, no. 4, pp. 207–211, Dec. 1994, access on http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1376556/pdf/jmedeth00293-0009.pdf

[13] Source: L. Alexander, “The treatment of shock from prolonged exposure to cold, especially in water”, Combined intelligence objectives subcommittee. Target no 24, report no 250. Washington, DC: Office of the Publication Board, Department of Commerce, 1946

6.2 Copyright

Exercise 10: Swiss Life - The Loto copyright compliance game

Answer the below questions with yes and no with the help of the 4 following statements. When you covered the plate, shout: carton!

1. "Let's remind you that commits plagiarism he who, even if he has obtained the copyright or if the work is in open access, does not cite his sources and thus lets the reader believe he is the author of a passage or an idea that he really has taken or adapted"[14]

2. "The free flow of ideas also needs to be unhindered. Ideas are thus not protected by copyrights; protection is limited to the form expressing an idea, such as the written text"[15]

3. A copyright transfer agreement is a legal document containing provisions for the conveyance of full or partial copyright from the rights owner to another party (…). Such agreements are a key element of subscription-based academic publishing"[16]

4. The intact or modified reproduction of a figure or text extract is authorized only for in-house educational purposes within an institution as long as sources are mentioned"[17]„

I can copy/paste and cite a published *graphic* of my own for				
I can copy/paste and cite a published *image* of my own for…				
I can copy/paste and cite a *book chapter* of my own for…				
I can copy/paste and cite a *published article* of my own for…				
	…my **PhD annual report?**	… my internal **teaching course notes**	… my **thesis?**	… my peer-reviewed **article's publication?**

[14] Source: Prof. Margaritondo, 20.01.2009, Flash, EPFL (freely translated by Chimène Glayre, 10.8.2010)
[15] Source: Copyrights, Swiss Federal Institute of Intellectual Property 06.02.2014, on https://www.ige.ch/en/copyright/copyrights.html
[16] Source: Copyright transfer agreement, Wikipedia, on http://en.wikipedia.org/wiki/Copyright_transfer_agreement , read the 31.03.2014
[17] Source : Michel Jaccard's presentation *Open Access Questions juridiques et meilleures pratiques,* in séminaire Droit d'auteur en bibliothèque, Forum 2009 bibliothèques HES-SO, 1er septembre 2009 – Neuchâtel

6.3 Creative Commons for easy reuse of information

The reuse of scientific information as is not allowed when the document is diffused outside the institution other than for internal in-house educational purposes because of copyright transfer agreement to publisher.

To reuse information as a figure (or long text):

- Ask the permission of reuse to copyright holder(s), publisher and/or author(s), for free or for fee
- Use similar figure of your own that was not published before
- Regenerate graph or table after asking the permission of reuse of data to author
- Redraw totally the figure
- Use similar figure under Creative Common license, by using Search Creative Commons[18] tool and check license compatibility

Miniguide of some Creative Commons licenses

	CC-BY	CC-BY-SA	CC-BY- ND	CC-BY-NC	CC-BY-NC-ND	CC-BY-NC-SA
Meaning	Paternity	Paternity Share-Alike	Paternity No derivative	Paternity Non commercial	Paternity Non commercial No derivative	Paternity Non commercial Share-Alike
You can	Copy, modify and diffuse with any CC license even for commercial purpose	Copy, modify and diffuse with any CC license even for commercial purpose	Copy and diffuse with any CC license even for commercial purpose	Copy, modify and diffuse for a non-commercial purpose	Copy and diffuse for non-commercial purpose	Copy, modify and diffuse for a non-commercial purpose
You have to	**Cite the author and the original license, and provide a link to the original source**					
		Publish the result under the same license	Publish the unchanged original source	Publish the result with non-commercial license	Publish the result with non-commercial license Publish the unchanged original source	Publish the result with non-commercial license Publish the result under the same license

[18] http://search.creativecommons.org/?lang=fr [accessed 27.03.2014]

7. Bibliometrics

What are Impact Factor (IF) and h-Index?

15h45	16h15	30’

Section objectives

1	The participant knows how to get or compute standard statistics of **IF** and **h-index**	
2	The participant is aware of common **pitfalls of IF** and **h-index**	

From search tools to scientometrics

Approximately six millions of documents published in 2012 were added to the Web of Science, a steady 5% increase over the past decade (**Figure 12**)

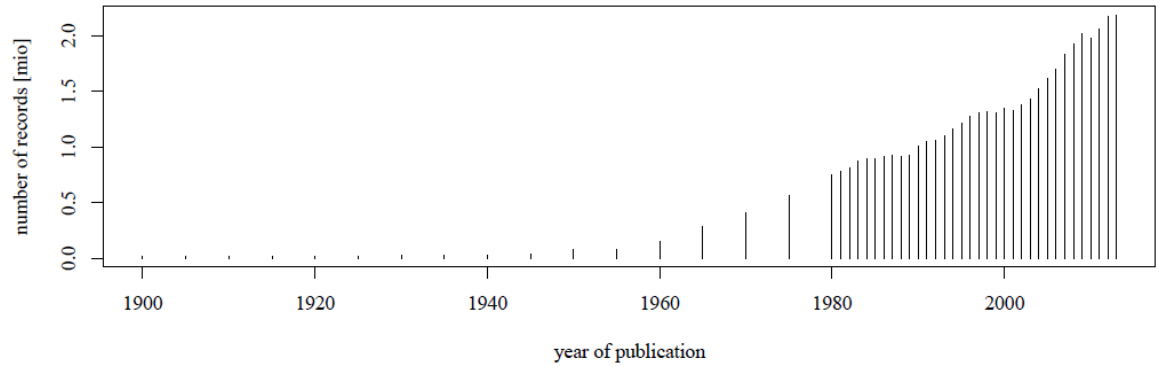


Figure 12 Increase of number of publications indexed by WOS[19]

Given this overwhelming amount of literature to read, it is tempting to rely on numerical summaries rather than digging into data in more detail. Bibliometrics, the activity of measuring scientific publication, evolved from a collection of tools used by libraries for internal administrative purposes, into scientometrics, a collection of widely used decision making tools supposed to help researchers in choosing journals, and universities in hiring researchers.

There are two kind of information that can be used as a proxy for so called calculation of scientific impact:

- The **number of publications** reflects a researcher's output. Someone having published a lot has made many contributions to science.
- The **number of citations** to a publication reflects the quality of that publication. A highly cited article is a good article.

Basically, any bibliographic database could count publications and how many times they are cited. But quality of data and built-in features vary greatly. The most widely used databases are Thomson Reuter's Web of Science (the oldest and de facto standard), Elsevier's Scopus and Google's Scholar. These three products are multidisciplinary tools that cover broad areas of science. Many other well established data bases exist that are more specialized, such as the National Institute of Health's PubMed, Elsevier's Engineering Village or the American Mathematical Society's MathSciNet, but only a few of them track citations.

As a quantitative method, bibliometrics pretends to be objective, hence scientific. However, how Thomson Reuters and Elsevier choose to include the journals they index (the initial and necessary step before any quantitative computation) seems to be based on purely qualitative and more or less trade secret criteria.

[19] Data and graph by Julien Junod, Bibliothèque de l'EPFL, December 2014

Bias of "machine" citation counting

Almost all bibliometric measures simply count items, whatever their content.

The two claims above make sense only if:

1. Every scientific publication would have been checked by a high quality peer review process for its quality, i.e. scientific novelty, soundness, etc.

2. Scientist would cite other contributions only because they think it is important, because they want to put their own work in perspective, etc (see chapter about citation), and avoid excessive autocitation.

3. Contribution of each author's (first, last, etc) would be equal

4. Each publication (articles, conference papers, posters, books...) would involve the same amount of work

5. If automatic counting of citation would be reliable. In fact, the quality of data and built-in features vary greatly. When indexation of journals by crawler is done, predatory journals are taken in account for citation counting; the way the databases are fed and maintained also play a role: procedures vary greatly between fully automatic treatment and partial human interventions and interoperability with other publisher databases may also be an issue.

6. Field communities were of equal size: community of researchers on skin cancer researchers is higher than those working on very fundamental, specialized and niche fields, such as ecophysiology of Alps grasshoppers.

7. Scientist would always praise others when citing them. Why does an article get cited? A controversial article may generate more buzz than praise.

8. Pre-XXIth century citations may not be properly covered in all databases.

9. Researcher would not adapt and optimize their behavior according to the criteria by which they are evaluated (competition situation as it is the case in science). The measurement disturbs the system, as it is well known from quantum mechanics!

The two most widely used measures are the impact factor (IF) and Hirsch's Index (h-index). Many other measures have been developed since, offering weightings, trying to correct potential biases, but these two are the simplest examples and show two different ways to perform computations.

7.1 Impact Factor

For a given year, the IF of a scientific journal is the mean number of citations of each published article in this journal by other journal indexed by the citing tool for the two last year period.

New journals that may be indexed from the first article by the citing tool will get an IF after 2 years period, if elected by Thomson Reuter. Some annual or irregular publications do not publish anything for a year, what is influencing on the calculation. The IF is always concerning a defined period of time. It is possible to calculate the IF on any period of time. The Journal of Citation Reports (JCR) includes also an IF of 5 years.

$$IF_n = \frac{citations_{n \to n-1} + citations_{n \to n-2}}{articles_{n-1} + articles_{n-2}}$$

Example: 2014 impact factor (IF) of a journal is

IF = A/B =

where:

A = the number of times that all items published in that journal in 2012 and 2013 were cited by indexed publications during 2008.

B = the total number of "citable items" published by that journal in 2012 and 2013. ("Citable items" for WoS and Scopus for this calculation are usually articles, reviews, proceedings, or notes; not editorials or letters to the editor, because they are not peer-reviewed).

Perhaps the easiest way, for an author, to interpret the impact factor is to think about it as the expected number of citations to a typical article in a given journal, during the year following publication. This may work as long as the citation count does not vary too much from year to year.

Note that the metric is defined at journal level, not at article level. It does not make any sense to speak of the impact factor of a single article hence, an author!

Bias, objections and anomalies of IF

A measure like the Impact Factor relies on an average, which, in order to represent of the center of the data points, requires this data to be tightly and evenly concentrated around this virtual center. This is not the case with citation counts. There are always a very few highly cited articles and very long tails of rarely or never cited articles.

*Example (**Figure 13**): According to the Journal of Citations Report (JCR), Nature had a 2008 impact factor of 31.43. This may suggest that a typical Nature article published that year could have expected to be cited more than 30 times during 2009. But the following figure shows that the, in this sense, "typical" articles form only a minority. Out of the more than 2000 articles published in 2008, roughly a quarter has never been cited so far, and only a third reaches the threshold of 30 citations*

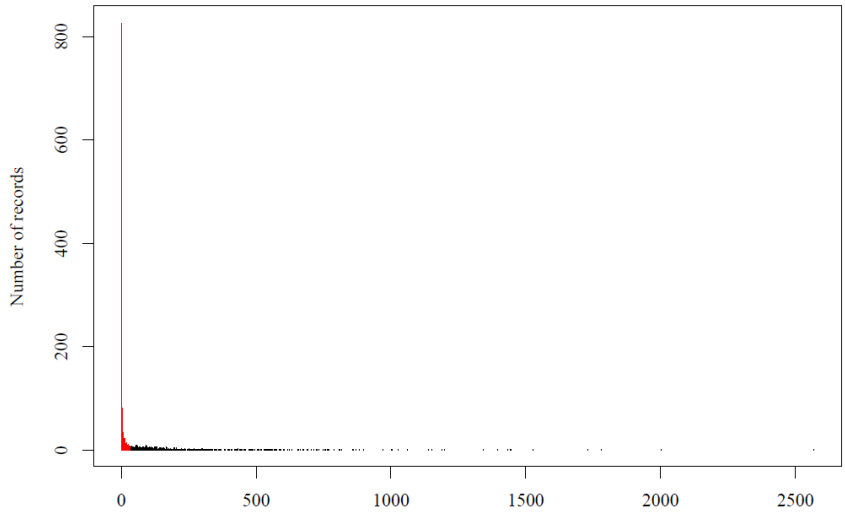


Figure 13 Distribution of WOS citations of articles published in Nature in 2008[20]

[20] Data and graph from Julien Junod, Bibliothèque de l'EPFL, December 2014

Exercise 11: IF anomalies

Find two IF anomalies in the chart and try to imagine possible explanations with a partner

(data from Journal of Citation Report edition 2015)

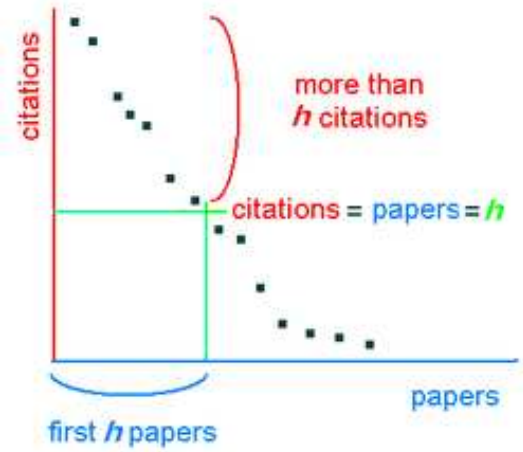
	Nature	***Science***	***Ca−A Cancer Journal for Clinicians***	***Acta Crystallographica Section A***
IF (2008)	*31.4*	*28.1*	*74.6*	*2.1*
IF (2010)	*36.1*	*31.4*	*94.3*	*54.3*
IF (2013)	*42.3*	*31.4*	*162*	*2.7*

Comment

7.2 h-index

In 2005, the physicist Jorge Hirsch proposed a more robust metric which aims at considering simultaneously the number of publications and their impact. An index of value of h says that h publications have been cited at least h times.

A scientist has index h if h of his or her Np papers have at least h citations each and the other (Np -h) papers have ≤ h citations each.



Example According to the Scopus database, on November the 20th 2014, Peter W. Higgs, who won the Nobel Prize in physics in 2013, wrote 11 documents indexed by Scopus. Ordering the publications by number of citations Nh known by Scopus, we can easily find his h-index:
the fifth ranked publication N_5 = 54: the number of citations is higher than the rank
the sixth ranked publication N_6 = 5: the number of citations is smaller than the rank
Therefore, Higg's h-index is 6.

N_p	**Publication title**	**Number of citations N_h**
1	Broken symmetries and the masses of gauge bosons	933
2	Broken symmetries, massless particles and gauge fields	837
3	Spontaneous symmetry breakdown without massless bosons	443
4	Dynamical symmetries in a spherical geometry	206
5	Quadratic lagrangians and general relativity	54
6	Integration of secondary constraints in quantized general relativity	5
7	Erratum: Integration of secondary constraints in quantized general relativity	2
8	Vacuum expectation values as sums over histories	1
9	A method for computing zero-point energies	1
10	Perturbation method for the calculation of molecular vibration frequencies	0
11	An application of perturbation theory to the F and G matrix method of calculating molecular vibration frequencies	0

Each database is likely to produce a different *h* for the same researcher, because of different coverage (**Table 9**)

Table 9 Total estimated indexed scholar journals

World	***WOS***	***Scopus***	***Google scholar***
40'000[21]	*18'000*[22]	*12'000*[23]	*Unknown, but every journal or conference technically compatible with Google Scholar crawlers are indexed*

- Web of Knowledge has strong coverage of journal publications till 1900, but poor coverage of high impact conferences. The exclusion of conference proceedings papers is a particular problem for scholars in computer science, where conference proceedings are considered an important part of the literature.
- Scopus has better coverage of conferences, but poor coverage of publications prior to 1996
- Google scholar has the best coverage of conferences and most scholar journals (though not all), but like Scopus has limited coverage of pre-1990 publications. Google scholar identifies about most of the time much more citations than Web of Knowledge and Scopus combined. Additional

[21] Source: Ulrichsweb.com
[22] Source: Journal of Citation Report (JCR) 2015
[23] Source: Scopus journal coverage 2015

citations reported by Google scholar may be explained by citations from low-impact journals, predatory journals, conference proceedings, grey literature, and much higher number of indexation of scholar journals and conferences (see above)

Bias, objections and anomalies of h-index

Fair comparisons imply similar data, but publication habits vary greatly among disciplines and over time. Therefore, comparisons across disciplines should be prohibited.

Example: All six of them won the Nobel Prize in 2013, but their publication statistics are not the same

Domain	***Name***	***Scopus h-index***
physics	*François Englert* *Peter W. Higgs*	*19* *5*
chemistry	*Martin Karplus* *Michael D. Levitt* *Arieh W. Warshel*	*117* *45* *89*
medicine	*James E. Rothman* *Randy W. Schekman Thomas Südhof*	*86* *79* *131*

Example: Nobel prize or equivalent researcher have a lower h-index than the most prolific fake data producer!

Author	***Prizes***	***Scopus number of publications***	***Google scholar number of publications***	***Scopus h-index***	***Google scholar h-index***
A. Einstein	*Nobel*	*18*	*210*	*0*	*98*
G. Perelman	*Fields Clay Millennium*	*2*	*31*	*0*	*13*
Y. Fuji	*World record of retractions*[24]	*212*	*n/a*	*39*	*n/a*

Some claims that Altmetrics will cure Bibliometrics from its statistical deceases. However, most of these "new" metrics seem to carry the same methodological flaws (e.g. blind counts) and simply work on different data. There have been enough reports about journals cheating with impact factors and zombie accounts on social networks to fear that it will be even easier to cheat with such web 2.0 metrics

[24] Source: Amarcus41 (02.074.2012) Does anesthesiology have a problem? Final version of report suggests Fujii will take retraction record, with 172. On http://retractionwatch.com/2012/07/02/does-anesthesiology-have-a-problem-final-version-of-report-suggests-fujii-will-take-retraction-record-with-172 [accessed 17.03.2015]

Miniguide of bibliometrics use

What to Do

- Always check the data behind high or low scores. Impressive values should immediately trigger the question “Why is that so?".
- Think about all the others informations you have. Do they corroborate or contradict the scores? Ask yourself if you did not already know what bibliometrics tell you.

What Not to Do

- Do not trust databases containing millions of records that have been fed automatically to be free of errors.
- Do not get hypnotized by all those numbers behind the period. Only differences in orders of magnitudes are really significant.
- Do not stick to absolute values. Measuring means comparing.
- Never compare scores accros disciplines.
- Never use Impact Factors for assessing articles or authors.

For your carrier survival

- Know your scores
- Know your competitor's scores
- Know if your scores are really a reasonable estimation of your scientific value
- Understand why this could not be the case. Be ready to argue!