

# Scholarly communication in the digital age

A practical introduction to Open Science and Research Data  
Management

**Francesca Di Donato**

 [orcid.org/0000-0003-0144-8934](https://orcid.org/0000-0003-0144-8934)

 @ederinita

[f.didonato@gmail.com](mailto:f.didonato@gmail.com)

## A few words about me

My name is **Francesca Di Donato**, and I'm researcher at the Institute of Computational Linguistics "A. Zampolli" at CNR.

I have a **PhD in History of Political Philosophy** and I worked at the University of Pisa, at the Scuola Normale Superiore and at Net7 srl. My research is focused on **Science Communication** since more than 15 years.

I'm involved in Eu projects (such as TRIPLE) and initiatives (such as CO-OPERAS Go-FAIR Implementation Network) on Open Science. I'm part of the ICDI Competence Center on EOSC.



Please, introduce yourself!

*Tell us a few words about you and your research topics*

(answers on mentimeter)

# What is this course about?

Question: *What are your expectations about this course?*

(answers on mentimeter)

# Outline of this course

1. Scholarly communication: an introduction. From the traditional paradigm to Open Science
2. The Open Access publishing paradigm and FAIR (Open) Data
3. Research Data Management and Data Management Plans
4. Research Infrastructures, the EOSC, OS skills and training, Science evaluation and OS workflows

1.

Scholarly communication:  
from the traditional paradigm to  
Open Science

# What is scholarly communication

“the system through which research and other scholarly writings are **created**, **evaluated** for quality, **disseminated** to the scholarly community, and **preserved** for future use” (ACRL Scholarly Communications Committee, 2006).

**Formalised publishing practices are just a subset of a larger pool of various communication practices** (emails, social media, blogs, press, etc.), both between scholars and between scholars and the public.

“includes both the dissemination and access to scholarship and research in a variety of formats and states of completion, such as published books or journal articles, research results and data sets, and drafts of papers” (Husain & Nazim, 2013, 405).

# What is scholarly communication

SC is the sum of common scholarly practices

..these practices are codified

..the advent of the Internet and the Web transformed the traditional communication system

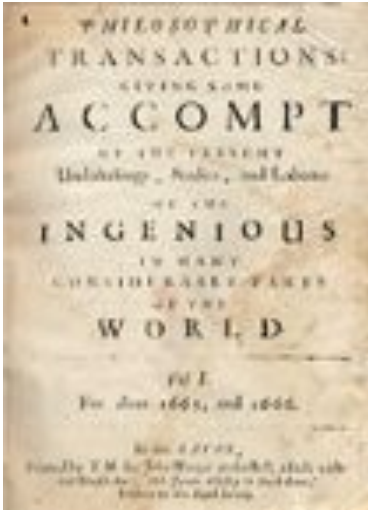
and nowadays, scholarship practices can be structured as open science practices, where Open Science is just science done right.

Thus, this course is about open science.



# Science communication: the modern system

## *London, XVII century*



### **Birth of Scientific Journals**

- 1665: *Philosophical Transactions of the Royal Society of London*
- Public register of Intellectual Property
- Parliament conferring Intellectual “Nobility”
- Arbiter elegantiarum

# Science communication: the modern system

## London, XVII century



## Birth of copyright (legal framework)

- 1710 Statute of Anne
- First copyright law (end of the system of privileges)
- 14 years (+14)

# Science communication: the modern system

*London, XVII century*



**Birth of peer reviewing (evaluation procedures)**

- Perusal (to get a fellowship)
- A work then can be printed or not

# After World War II

In 1960s, Eugene Garfield (ISI) creates the Science Citation Index (now Web of Science): a bibliographic instrument born to provide a “Cartography of Citations”. It was conceived to **select the most popular Journals in a given field** that a library should buy.

## *Side-effects*

“**Impact Factor**”: a standardized measurement form introduced by the ISI that allows to determine the impact of an article on later publications;

“**Core Journals**”: based on a notion of “fundamental Journals for a fundamental Science”;

**Journals’ functions**: they grant intellectual property rights; they provide a brand; they work as career management system.

## H-index

For a **researcher**: the maximum value of  $h$  such that the given author has published  $h$  papers that have each been cited at least  $h$  times

For a given year  $y$ , the specific **journal** Impact Factor is the number of citations, received in that year, of articles published in that journal during the two preceding years, divided by the total number of "citable items" published in that journal during the two preceding years

## Impact Factor

Source: slide n. 24, E. Lazzeri, F. Di Donato, [Open Science: Why it is important](https://zenodo.org/record/4317277), 10.5281/zenodo.4317277

# Citation based indexes criticism

- Early career researchers are penalised
- The citation context is not considered (eg. Negative citation)
- They are influenced by the limitation of the citational databases (which are all owned by big scientific publishers)
- It can be manipulated by both authors and reviewers (self and cross citations)
- It does not take into account the number of authors in a paper and their contribution given
- It does not take into account research multidisciplinary
- It does not facilitate science freedom.

DORA, 2013, <https://sfdora.org/read/>

McKiernan, et al, 2019. <https://elifesciences.org/articles/47338>

Niles, et al, 2019. <https://www.biorxiv.org/content/10.1101/706622v1>

Alder, et al, 2008. <https://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>

Source: slide n. 25, E. Lazzeri, F. Di Donato, [Open Science: Why it is important](https://zenodo.org/record/4317277), 10.5281/zenodo.4317277

# Impact Factor: Criticisms

- Most recent Journals – even if prestigious – are not assigned an IF for many years.
- Eugene Garfield and Clarivate Analytics warns about using JIF to evaluate researchers because the quality of single papers in a Journal cannot be judge on the basis of a medium index.
- It is discipline dependent and strongly affected by mainstream topics
- It has never been certified by independent studies (data are not publicly available)

# A “publish or perish” system

Scholars are forced to publish their results (for grants, careers, fundings)

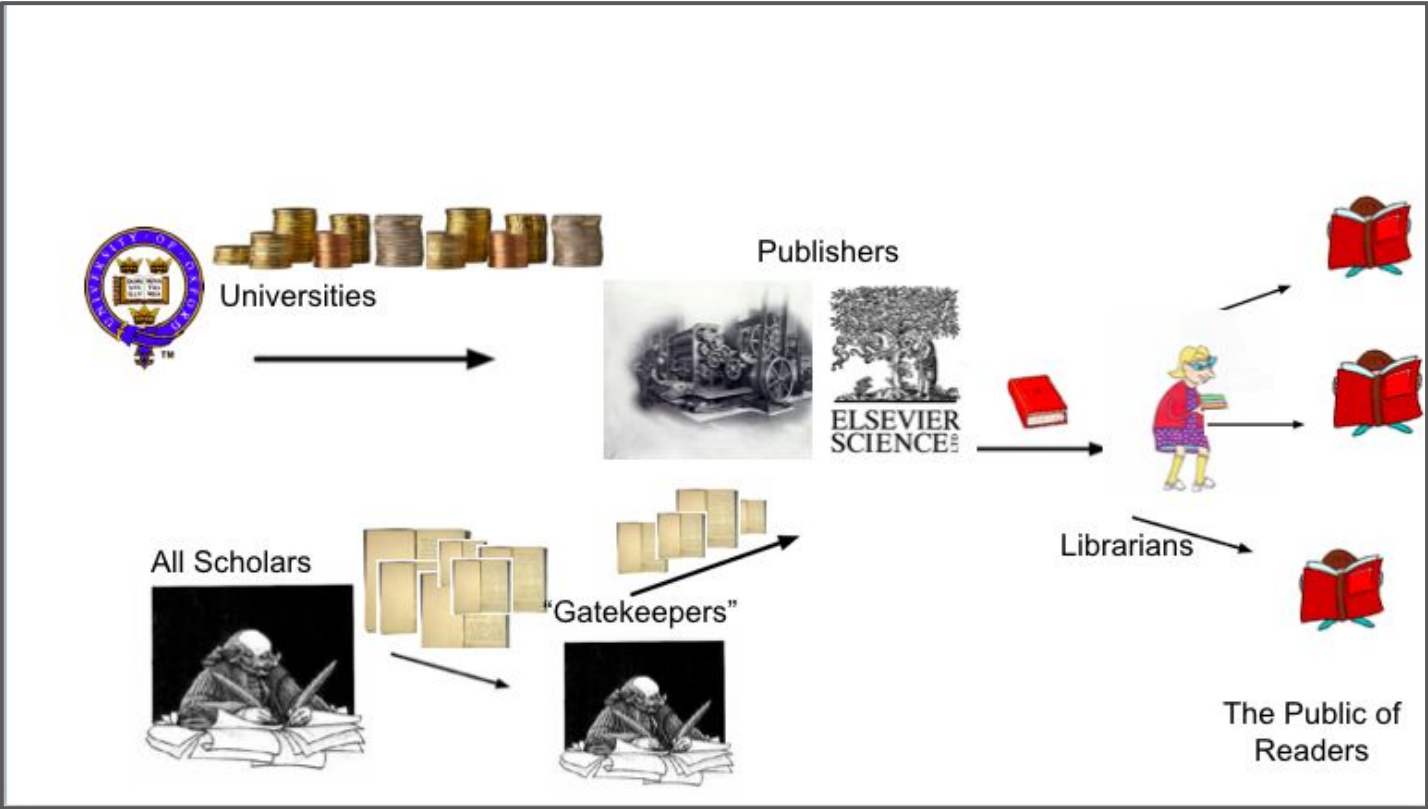
The practice of “salami publishing”

Quantity prevails on quality

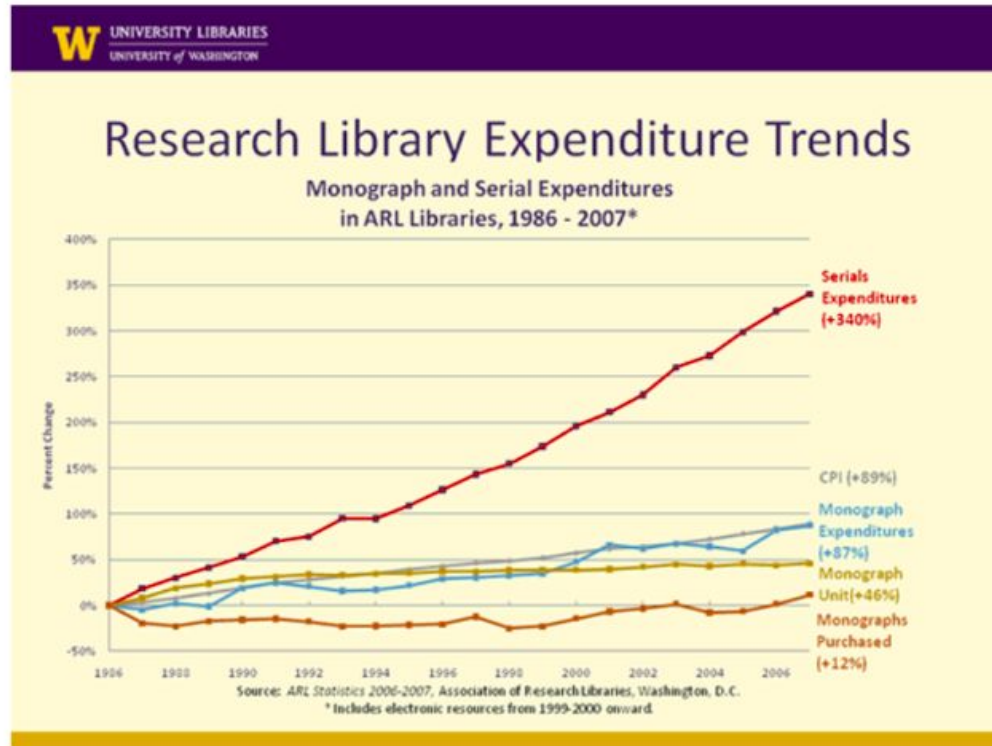
Inelastic market → Serial Price Crisis



# Funding mechanism in the traditional system



# The Serial Crisis



<http://www.librarytechnology.org/automationhistory.pl>

# The “open access” movement

## ***Practices***

- P. Ginsparg: (xxx) ArXiv.org Los Alamos, 1991
- PubMed, PLOS

## ***Technologies***

- Open Archives Initiative, Santa Fe, 1999
- OAI-PMH v. 1.0, 2001 e v. 2.0, 2002

## ***Policies***

- Budapest Open Access Initiative, 2002
- Bethesda Statement on Open Access Publishing, 2003
- Berlin Declaration to Open Access, 2003

**“An old tradition  
and a new technology  
have converged  
to make possible  
an unprecedented public good”**

([Budapest Open Access Initiative](#) (BOAI), 2002)

# The Advent of the Internet and the Web

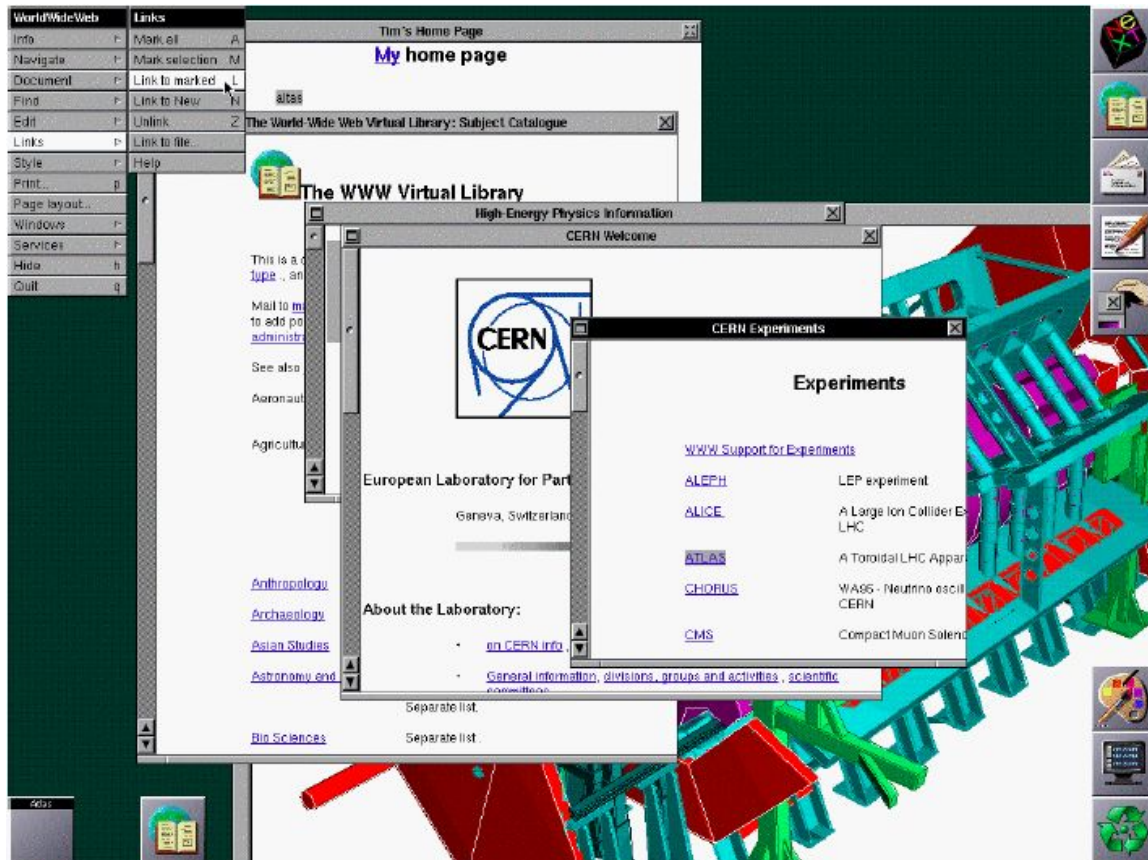
Question: *What Is the Internet? What the Web? What's the difference between them?*

(answers on mentimeter)



# The Web as a research environment

# The birth of the Web





On the shoulders of giants

# Internet

## Architecture

distributed

flexible

open

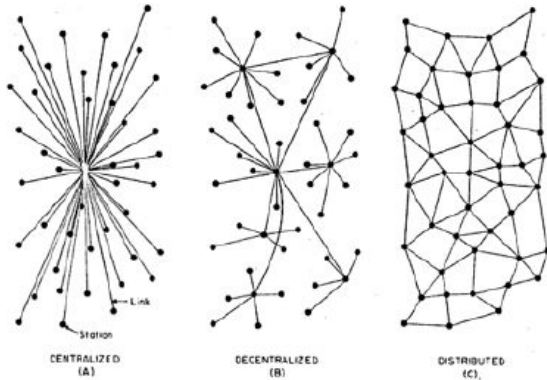
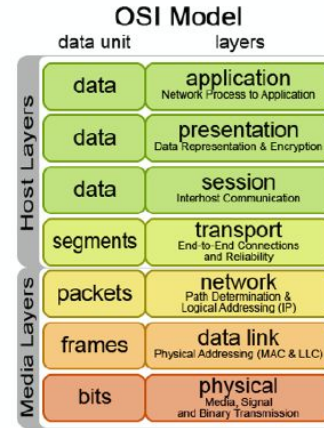


FIG. 1 - Centralized, Decentralized and Distributed Networks



scalable



# The Web (1989)

## 3 principles

URL

HTTP

HTML

Adoption of **standards** for:  
chars encoding (i18n)  
accessibility

**Independence from**  
hardware  
software  
scale

**Separation** between  
data and  
documents

# Web Architecture

The Web is a global hypertext/hypermedia based on 3 principles and technologies:

1. URI : Universal Resource Identifier (ex. An URL)
2. HTTP: HyperText Transfer Protocol (on TCP/IP)
3. HTML: Hypertext Markup Language

# A universal information system

“Suppose all the information stored on computers everywhere were linked, I thought. Suppose I could program my computer to create a space in which anything could be linked to anything. All the bits of information in every computer at CERN, and on the planet, would be available to me and to anyone else. There would be a single, global information space.”

The web is your research library:  
What about its topology?

# Topology of the Web



(Koenigsberg bridges, 1736)

**See:**

A.L. Barabasi,

[Linked. The new Science of Networks](#), 2002.

The web is a **graph**: an abstract representation of a set of objects where some pairs of the objects are connected by links.

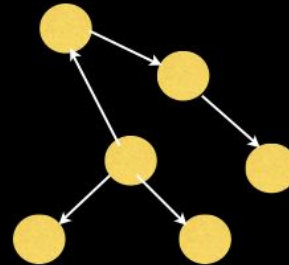
The interconnected objects are represented by mathematical abstractions called **vertices**, and the links that connect some pairs of vertices are called **edges**.

Typically, a graph is depicted in diagrammatic form as a set of dots for the vertices, joined by lines or curves for the edges.

# Web topology

The edges may be directed (asymmetric) or undirected (symmetric).

The web is a **directed graph** (links go in only one direction).



like the scholarly publications graph (where nodes are papers and links are citations)



Plato



Kant



Di Donato

**See:**

A.L. Barabasi,  
[Linked. The new Science of Networks](#), 2002.



**See:**

A.L. Barabasi,  
[Linked. The new Science of Networks](#), 2002.

## The Web is a **small world network**:

a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of steps.

A small world network, where nodes represent people and edges connect people that know each other, captures the small world phenomenon of strangers being linked by a mutual acquaintance.

**See:**

A.L. Barabasi,  
[Linked. The new Science of Networks](#), 2002.

## Six degrees of separation:

if a person is one step away from each person they know and two steps away from each person who is known by one of the people they know, then everyone is at most six steps away from any other person on Earth.

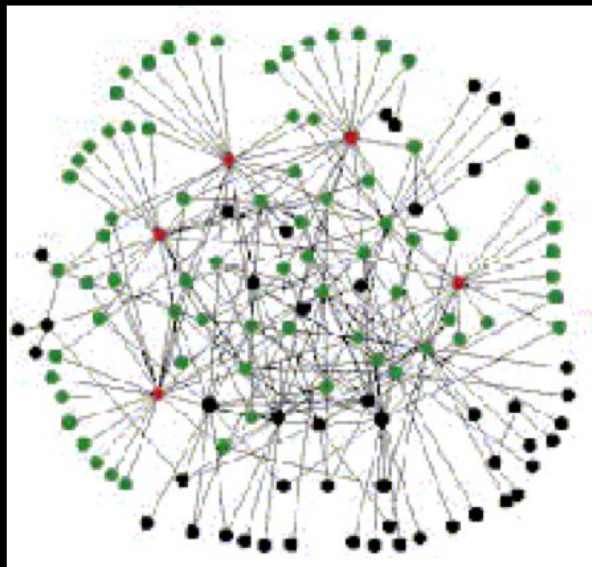
In 2004, the degrees of separation on the Web were 19.

**Ex.** [http://en.wikipedia.org/wiki/Six\\_degrees\\_of\\_Kevin\\_Bacon](http://en.wikipedia.org/wiki/Six_degrees_of_Kevin_Bacon)

# On the Web, not all the nodes are equal: there are **hubs** and **authorities**

**See:**

A.L. Barabasi,  
[Linked. The new Science of Networks](#), 2002.



<http://www.nd.edu/~networks/Linked/>

A scale-free network of 130 nodes generated by the scale-free model.

The five biggest nodes are shown in red, and they are in contact with 60% of nodes (green).

# The Pareto (or 80/20) principle

**See:**

A.L. Barabasi,  
[Linked. The new Science  
of Networks](#), 2002.

The Pareto principle (also known as the 80-20 rule) states that, for many events, roughly 80% of the effects come from 20% of the causes.

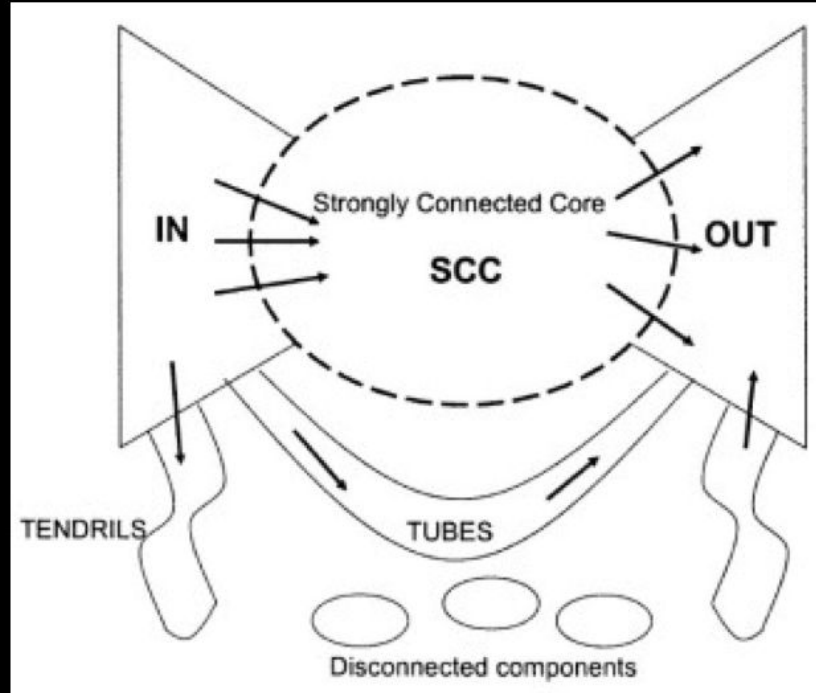
The 80% of Web links come from 20% of Web pages.

It has 4 continents.. but we can explore only two of them

**See:**

A.L. Barabasi,

[Linked. The new Science of Networks](#), 2002.



<http://thenextweb.org/wp-content/uploads/2008/06/fragmented.jpg>

Source: Slide 36 of F. Di Donato, [The Web as a Research Environment. Definitions, tools, and best practices.](#)

**See:**

A.L. Barabasi,

[Linked. The new Science of Networks](#), 2002.

Note that  
the scholarly  
publications  
network  
obeys  
to the same rules





# Quality: What about evaluation on the Web?

Many documentation systems used to be designed for particular collections of information, and one could assume that the information in such a system had achieved a certain quality. However, **the Web itself cannot enforce any single notion of quality. Such notions are very subjective, and change with time.** To support this -- to allow users to actually use the web even though it contains junk as well as gems -- the technology must allow powerful filtering tools which, combining opinions and information about information from many sources, are completely under the control of the user.

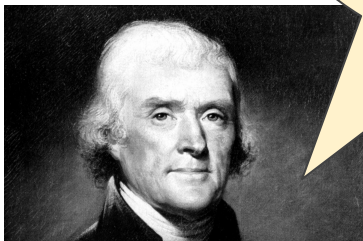
It is understood that a collection of works, such as a set of technical reports or a library, only includes articles reaching a certain standard, and some early dial-up information services similarly amassed information according to some quality criterion. Some people miss that with the Web -- hence the need for portals which provide a filtered view. However useful people find such portals, though, **it is important that the Web itself doesn't try to promote a single notion of quality.**

T. Berners-Lee, "The World Wide Web - Past, present, future. Exploring Universality", 2002. <https://www.w3.org/2002/04/Japan/Lecture.html>



“An old tradition and a new technology have converged to make possible an unprecedented public good” (BOAI, 2002)

He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me.



**Thomas Jefferson** (1813)

...but also **Giustino filosofo**: «Nonne vero hoc tale est quale etiam in nobis fieri videmus? Sermonen enim aliquem proferentes, sermonem gignimus; non per abscissionem, ita ut sermo (ratio) qui in nobis est imminuatur, proferentes. Quale est etiam quod in igni videmus alium [ignem] fieri, non imminuto illo ex quo accensus est, sed in eodem statu manente; et qui ex eo accensus est etiam ipse existens apparet non imminuens illum ex quo accensus est»

If I have seen further it is by standing on the shoulders of Giants



**Isaac Newton** (1676)

..but also **Giovanni di Salisbry**: «Bernard of Chartres used to say that we [the Moderns] are like dwarves perched on the shoulders of giants [the Ancients], and thus we are able to see more and farther than the latter. And this is not at all because of the acuteness of our sight or the stature of our body, but because we are carried aloft and elevated by the magnitude of the giants.»

# Galileo Galilei, letter to Belisario Vinta, about the Sidereus Nuncius (1610)

**"It seems necessary to me, over and above other precautions, to maintain and increase the popularity of these discoveries, that as many people as possible see and recognize this truth."**

- Openness = Publicity of knowledge (through the press)
- Both oriented to experts and non-experts (science as a common good)

# The “open access” movement

## ***Practices***

- P. Ginsparg: ArXiv.org, Los Alamos, 1991
- PubMed, PLOS

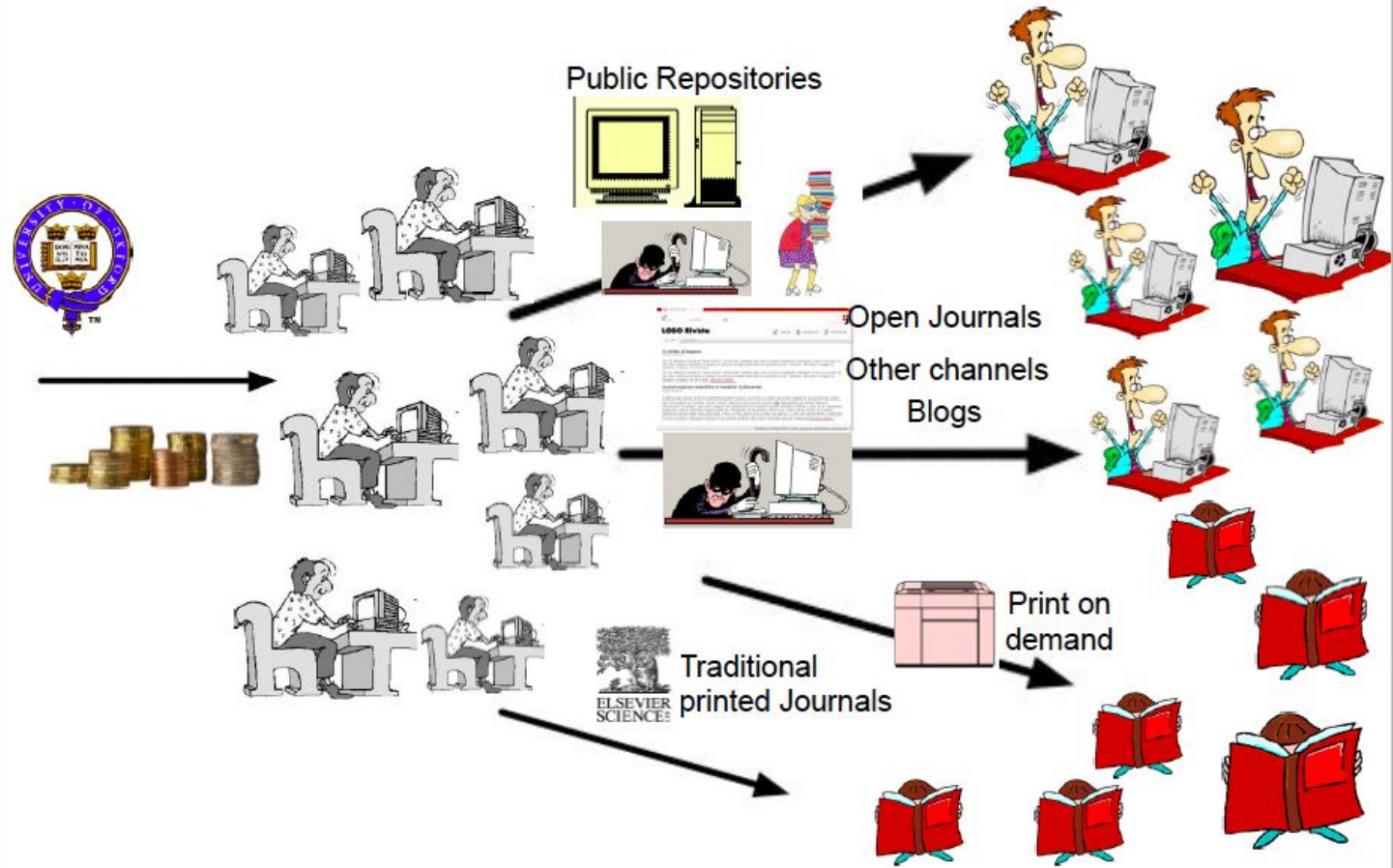
## ***Technologies***

- Open Archives Initiative, Santa Fe, 1999
- OAI-PMH v. 1.0, 2001 e v. 2.0, 2002

## ***Policies***

- Budapest Open Access Initiative, 2002
- Bethesda Statement on Open Access Publishing, 2003
- Berlin Declaration to Open Access, 2003

# A new scenario



# Open Access

Content and software tools openly accessible and compatible.

Content: Public-funded scientific results, that authors publish for free, such as original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.

"By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint .. should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited" (Berlin Declaration for Open Access, 2003).

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Oct. 2003)

<<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>>

- Budapest Open Access Initiative (2002-2004)

<<http://www.soros.org/openaccess/>>

# Ilaria Capua and Open Access to research data during the avian flu

nature > correspondence > article

## nature

Correspondence | Published: 30 August 2006

### A global initiative on sharing avian flu data

Peter Bogner, Ilaria Capua, David J. Lipman, Nancy J. Cox & others

*Nature* 442, 981(2006) | [Cite this article](#)

469 Accesses | 66 Citations | 24 Altmetric | [Metrics](#)



In 2006 Ilaria Capua and other scientists called for open access to research data related to the avian flu to collaborate on a global level to collectively study the new virus

Source: slide n. 49, E. Lazzeri, F. Di Donato, [Open Science: Why it is important](https://doi.org/10.5281/zenodo.4317277), 10.5281/zenodo.4317277

# Open Science =

---

- Open **source** - from 1987...
  - + WWWeb - from 1989...
- + Open **content** - from 1999...
  - + Web 2.0 - from 2004...
- + Open **access** - from 2003...
- + Open **data** - from 2006...
  - + Social
  - + Semantic Web? - now on...



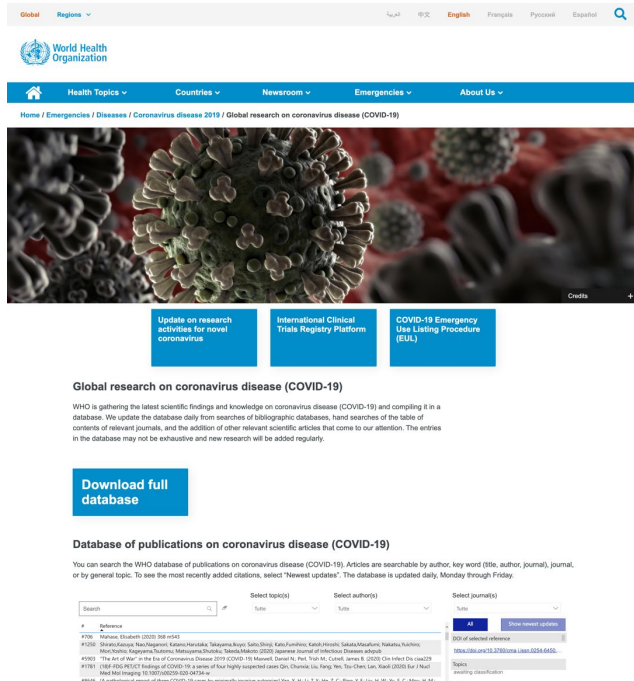
# how does science work (in practice)



<http://www.aib.it/aib/sezioni/emr/bibtime/num-xix-1/giglia.htm>



# Collaboration!



The screenshot shows the WHO website's dedicated page for COVID-19. At the top, there are navigation menus for 'Global', 'Regions', and various languages. Below this is a blue header with 'World Health Organization' and a search bar. The main content area features a large image of a coronavirus particle. Below the image are three blue buttons: 'Update on research activities for novel coronavirus', 'International Clinical Trials Registry Platform', and 'COVID-19 Emergency Use Listing Procedure (EUL)'. The page is divided into three main sections: 'Global research on coronavirus disease (COVID-19)', 'Download full database', and 'Database of publications on coronavirus disease (COVID-19)'. The 'Global research' section includes a paragraph about WHO's efforts to gather scientific findings and a 'Download full database' button. The 'Database of publications' section provides a search interface with fields for 'Search', 'Select topic(s)', 'Select author(s)', and 'Select journal(s)', along with a 'Download results' button.

[WHO COVID-19 dedicated page](#)

“I call on all countries, companies and research institutions to support open data, open science, and open collaboration so that all people can enjoy the benefits of science and research” (Ghebreyesus TA: ‘WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 6 April 2020’. 6 April 2020)

«It is imperative to leverage scientific innovations and support principles of openness and inclusiveness in processes that generate solutions to the severe health menace that is likely to bring significant hardships to humanity.»  
[UNESCO - Open access to facilitate research and information on COVID-19](#)



CBC.CA

'We're opening everything': Scientists share coronavirus data in unprecedented way to contain, treat disease | CBC News



The screenshot shows the Elsevier website's COVID-19 information center. The page features the Elsevier logo at the top right and a navigation menu. The main content area includes a headline: 'Elsevier gives full access to its content on its COVID-19 Information Center for PubMed Central and other public health databases to accelerate fight against coronavirus'. Below the headline are social media sharing icons and the text 'New York, March 13, 2020'.

Source: slide n. 50, E. Lazzeri, F. Di Donato, [Open Science: Why it is important](#), 10.5281/zenodo.4317277, slightly modified

# Open Science: Science the correct way!

- **Open Science** means: **open** each step of the research cycle.
- Open Science principles are: **transparency**, **reproducibility**, **collaboration**, **inclusiveness**, **accessibility**, **accuracy**, **re-use**
- Open Science steps from the concept that the research that is funded with **public money** has to be made **immediately available** to the community: «*every EU citizen has the right to access and benefit from knowledge produced using public funds*» [Neelie Kroes, European Commission]
- The **European Commission** and a long list of **International Funders** made a clear choice towards Open Science

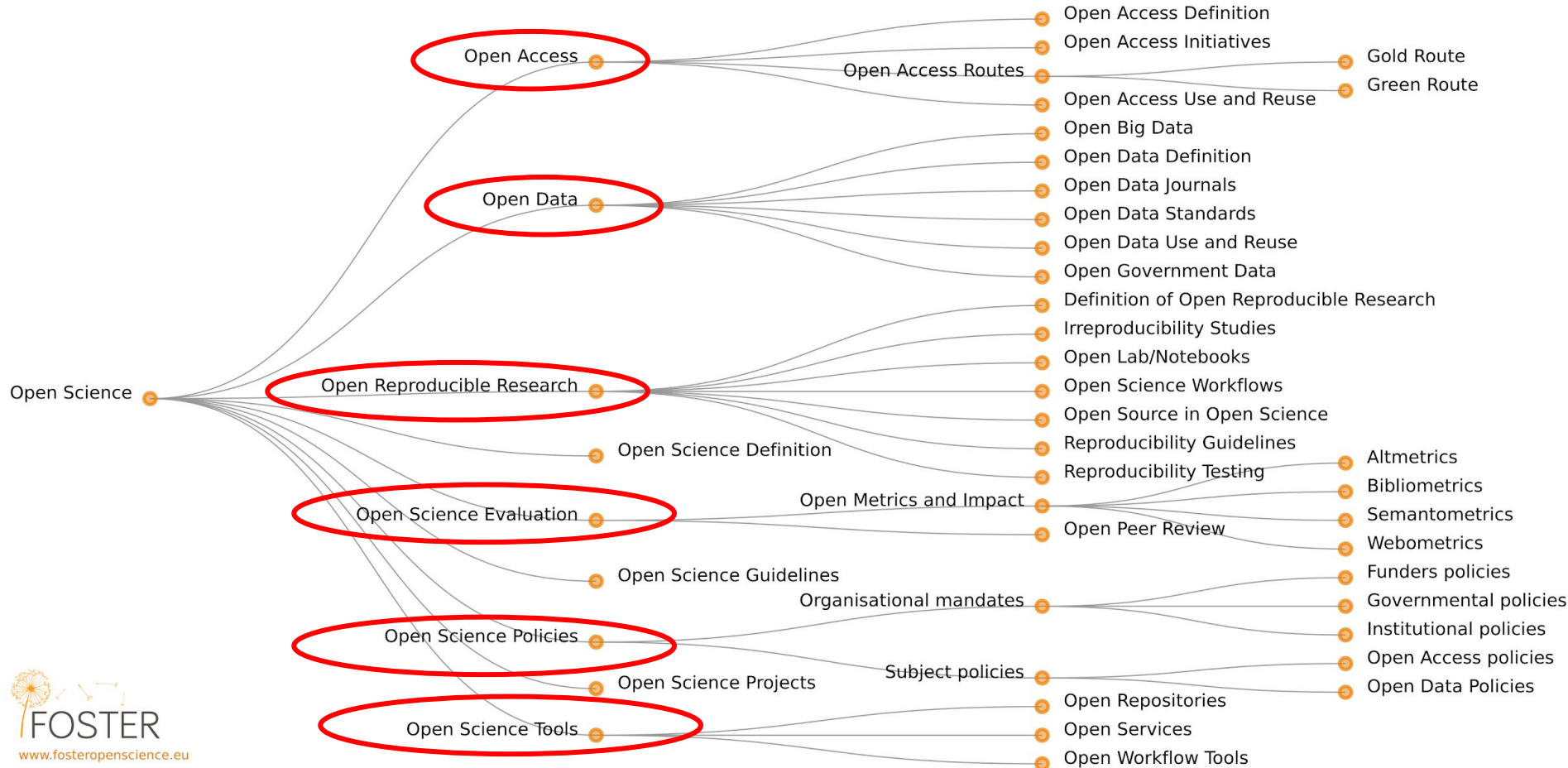
# Open Science

Open Science means a **broader access** to **publicly funded research results** and therefore helps to:

- build on previous research results (improved quality of results)
- encourage **collaboration** and avoid duplication of effort (**greater efficiency**)
- speed up **innovation** (faster progress to market means faster growth)
- **involve citizens** and society (improved transparency of the scientific process).

What we will cover in this course

# Open Science Taxonomy



# In the next lessons

Open access

Open/FAIR data (and research data management) → Open reproducible research

Open science evaluation

Open science policies

Open science workflows

Open science skills and training

Open science tools

- + The European context:
  - + Research Infrastructures
  - + The European Open Science cloud



questions?

(answers on mentimeter)



# References

ACRL Scholarly Communications Committee: Principles and Strategies for the Reform of Scholarly Communication. Text. Association of College & Research Libraries (ACRL). 2006. <http://www.ala.org/acrl/publications/whitepapers/principlesstrategies>

S. Husain, Nazim M, Analysis of Open Access Scholarly Journals in Media & Communication. *DESIDOC Journal of Library and Information Technology*. 2013; **33**(5): 405–11. <https://doi.org/10.14429/djlit.33.5106>

R. Alder, et al, Citation Statistics. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS, Joint Committee on Quantitative Assessment of Research, 2008. <https://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>

A.L. Barabasi, [Linked. The new Science of Networks](#), 2002.

T. Berners-Lee, "The World Wide Web - Past, present, future. Exploring Universality", 2002. <https://www.w3.org/2002/04/Japan/Lecture.html>

T. Berners Lee, *Weaving the web: The Original Design and Ultimate Destiny of the World Wide*, 2000.

P. Bogner, Capua, I., Lipman, D. *et al*. A global initiative on sharing avian flu data. *Nature* 442, 981 (2006). <https://doi.org/10.1038/442981a>

F. Di Donato, [La scienza e la rete](#). L'uso pubblico della ragione nell'età del web, FUP, 2009.

F. Di Donato, [The Web as a Research Environment. Definitions, tools, and best practices](#), 2016.

# References

P. Greco, [Il Sidereus Nuncius e l'origine della comunicazione pubblica della scienza](#), S&F\_scienzaefilosofia.it, 2018

J.-C. Guédon, [In Oldenburg's long shadow](#), ARL, 2001.

T. Jefferson, [Lettera a Isaac McPherson](#) del 13 agosto 1813.

E. Lazzeri, F. Di Donato, [Open Science: Why it is important](#), 10.5281/zenodo.4317277

E. McKiernan, et al, Meta-Research: Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations, 2019. <https://elifesciences.org/articles/47338>

M.T. Niles, et al, Why we publish where we do: Faculty publishing values and their relationship to review, promotion and tenure expectations, 2019. <https://www.biorxiv.org/content/10.1101/706622v1>

M.C. Pievatolo, [I padroni del discorso. Platone e la libertà della conoscenza](#), Methexis - PLUS, Pisa, p. 25.

## Links

- [Bethesda Statement on Open Access Publishing](#), 2003,
- [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#) (Oct. 2003),
- [Budapest Open Access Initiative](#) (BOAI), 2002
- [The Clarivate Analytics Impact Factor](#),
- [DORA](#), 2013,
- [Isaac Newton](#), on wikipedia.
- [Nani sulle spalle dei giganti](#), voce Enciclopedia Treccani.
- [Open Science Taxonomy](#)
- Research Library Expenditure Trends - [The Library of Tomorrow - thoughts and reflections on the world of Library and Information Science](#)
- [WHO COVID-19 dedicated page](#)

## Credits

- Slides coffee: Photo by [Mukul Wadhwa](#) on [Unsplash](#)
- Image of the slide 39 from: <https://medium.com/q-e-d/whos-important-a-tale-from-wikipedia-a370dc6ef078>

# Appendix 1

## Searching the Web

# How do we search the Web?\*

## (0) Exploring the Web surface: on the use of SES

“Though hundreds of search engines are freely and publicly available, a very few capture the overwhelming majority of the audience. According to the well-known 80/20 rule, **80 percent of users are concentrated on 20 percent of applications.**

**Users trust their own ability as web searchers.**

More than 90 percent of people who use search engines say they are confident in the answers; half are very confident. Users also judge their research activities as successful in most cases.

**The less Internet experience people have, the more successful they regard their own searches.”**

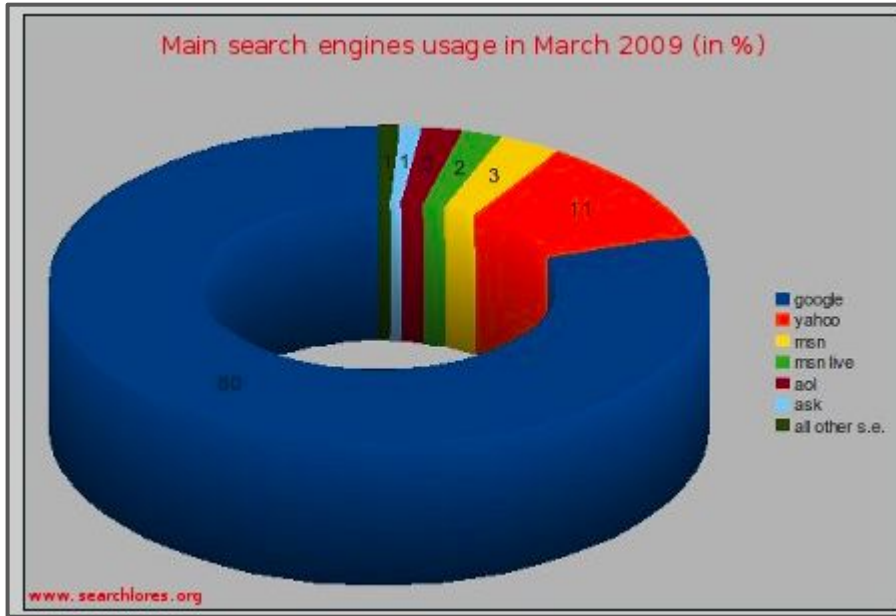
## In SES we trust!

“Students tend to place supreme confidence in the results of search engines, without having any idea how they work or being aware that they are fundamentally commercial operations.”

### But SES are many (many!) and different

“Search engines use in fact quite different algos, which gives indexes that do not overlap that much and thus offer searchers the possibility to fish results that they wouldn't even see if they would stuck onto just one index.”

On wikipedia you find an updated [List of search engines](#), divided by content/topic, by data type, by modes, by popularity, and venue type.



Fravia's website [Searchlores](http://www.searchlores.org), provides a lot of knowledge, useful tips, and search strategies.



Source: Fravia, [Searchlores - Main search engines](http://www.searchlores.org)

# You have to find relevant sources, such as:

1. Books: references (metadata) but also full texts
2. Articles: references (metadata) but also full texts
3. Conference proceedings: references (metadata) but also full texts
4. Research blogs
5. Events
6. Presentations: references (metadata) but also full texts
7. Various types of (open) research data
8. ....
9. ...

# How do we **search** the Web?

1. Search Engines
2. Specialized DB: such as G Scholar, Mendeley, Paperity, SSRN\*, Gallica\*\*, Web of Science, Scopus, JSTOR, and.....
3. Research blogs (Hypotheses\*\*,)
4. Social Networks: via twitter, via fb
5. Events
6. Open Access Directories
7. Various types of (open) research data: Registry of Research data Repositories\*, Open Data research Network\*\*

\* <http://www.ssrn.com/en/>

\*\* <http://hypotheses.org/>

\*\*\* <http://www.re3data.org/>

\*\*\*\* <http://www.opendataresearch.org/>

\*\*\*\*\* see also <https://www.rd-alliance.org/>



# How do we *get access*?

Via:

1. Your Institutional library/Pisa Library Network: get info about the subscriptions of your library, ask for books and articles they can get for you via their networks (go also to the SNS library, and the University of Pisa one. See the Metaopac pisano\* catalogue)
2. Open Access Journals, Books and OA Archives: DOAJ (<https://doaj.org/>), DOAB (<https://www.doabooks.org/>), OpenDOAR (<http://www.opendoar.org/>), ROAR (<http://roar.eprints.org/>).
3. Social Networks, research blogs, etc.: twitter, Research Gate / Academia to directly contact authors, via @email to authors

\*<http://leonardo.isti.cnr.it/p8090eduservice/servlet/lisis?Conf=/export/home/metaopac/mop2Conf/AIXnotlogic0.pisa.sys.file>

# The deep web

refers to World Wide Web content that is not part of the surface Web, which is indexed by standard search engines.

The deep Web is several orders of magnitude larger than the surface Web (in September 2001 the deep web was estimated to be 400 to 550 times larger than the surface web).

**Digital libraries** are a prime example of the deep web.

“A digital library is a focused collection of digital objects, including text, audio, and video, along with methods for access and retrieval, and for selection, organization, and maintenance of the collection”. I.H. Witten, [Browsing around a Digital Library](#)

Other examples?

## The deep web

Who's downloading pirated papers?

**EVERYON**

In rich and poor countries, researchers turn to the Sci-Hub website



# Online DBs for SSH

<http://www.gutenberg.org/>: Project Gutenberg

<http://gutenberg.net.au/>: Gutenberg australia

(As the oz law is just 50 years max...)

<http://gallica.bnf.fr/>: Gallica

<http://about.eserver.org/>: Eserver

<http://books.google.com/books?/>: Google books

<http://scholar.google.com/>: Google scholar

<http://en.scientificcommons.org/>: Index of OAI-compliant papers

<http://www.archive.org/search.php?query=subject%3A%22search%22>: Internet Archive

<http://vlib.org.uk/>: The WWW Virtual Library

<http://digital.library.upenn.edu/books/search.html>: The University of Pennsylvania Online Books Page

<http://abu.cnam.fr/index.html>: ABU: la Bibliothèque Universelle

<http://www.opencontentalliance.org/>: Open Content Alliance

<http://www.readprint.com/>: Our website offers thousands of free books for students, teachers, and the classic enthusiast.

<http://www.gutenberg.org/>: There are 17,000 free books in the Project Gutenberg Online Book Catalog.

<http://www.bibliomania.com/>: Free Online Literature with more than 2000 Classic Texts

<http://digital.library.upenn.edu/books/>: Upenn.edu, Listing over 25,000 free books on the Web

<http://www.ipl.org/div/subject/browse/hum60.60.00/>: The Internet Public Library, Literature Online Texts

## Classicalia

<http://www.thelatinlibrary.com/index.html>: the latin library, latin

<http://www.thelatinlibrary.com/neo.html>: the latin library, neo-latin

[http://www.forumromanum.org/literature/authors\\_a.html](http://www.forumromanum.org/literature/authors_a.html): Corpus scriptorum latinorum

[http://www.fh-augsburg.de/~harsch/a\\_chron.html#latmed](http://www.fh-augsburg.de/~harsch/a_chron.html#latmed): BIBLIOTHECA AUGUSTANA  
(Bibliotheca Latina scriptorum latinorum collectio)

<http://www.corpusthomicum.org/iopera.html>: CORPUS THOMISTICUM S. THOMAE DE  
AQUINO OPERA OMNIA

<http://www.textkit.com/>: Textkit is the Internet's largest provider of free and fully  
downloadable Greek and Latin grammars and readers. With currently 146 free books to  
choose from.

[http://www.molfettanet.com/tradizioni/pesi\\_e\\_misure.htm](http://www.molfettanet.com/tradizioni/pesi_e_misure.htm): Pesi e misure nell'antichità  
(Italian)

<http://www.ancientlibrary.com/wcd/>: The wiki classical dictionary at ancientlibrary (currently  
down :-{

<http://www.perseus.tufts.edu/hopper/>: Perseus Digital Library

## De re orientalia

<http://omphaloskepsis.com/collection/index.html>: Omphaloskepsis

Omphaloskepsis