



**HAL**  
open science

## A to Z, An Open Science Alphabet for Academic Librarians

Joachim Schöpfel, Hélène Prost

► **To cite this version:**

Joachim Schöpfel, Hélène Prost. A to Z, An Open Science Alphabet for Academic Librarians. ITlib, 2019, 4, pp.61-66. hal-02448603

**HAL Id: hal-02448603**

**<https://hal.science/hal-02448603v1>**

Submitted on 13 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A-to-Z, An Open Science Alphabet for Academic Librarians

Joachim Schöpfel (University of Lille, GERiiCO laboratory)

Hélène Prost (CNRS, GERiiCO laboratory)

## Abstract

The paper provides an informed and up-to-date description of some central terms useful for a better understanding of the ongoing projects and debates of the global open access and open science movement.

## Key words

Open science, open access, open data

## Amsterdam Call for Action

Based on the input of experts and stakeholders of the Amsterdam Conference “Open Science – From Vision to Action”, hosted by the Netherlands’ EU Presidency on 4 and 5 April 2016. The Call stated two goals for 2020: full open access for all scientific publications, and a new approach towards optimal reuse of research data. As flanking measures, it promoted new assessment, reward and evaluation systems, alignment of policies and exchange of best practices. Publishers regarded the Amsterdam Call as successful lobbying because they are mentioned 25 times in the document, as essential stakeholders in 9 out of 12 concrete actions.

## APC

Article processing charges. Key element of many open access business models. The amount authors or their institutions must pay to some journals to get their papers published in immediate open access. According to OpenAPC, they vary between €232 and €5,104, with a median of €1,640 (October 2019). Are correlated with impact factor and the journals’ reputation. Should have introduced more competition into the journal market and limited the price increase; so far, have not.

## Berlin Declaration

Second major declaration 2003 (after Budapest 2002) in favor of open access to scientific results and one of the milestones of the open access movement. Was written by and for research policy makers, research institutions, funding agencies, libraries, archives and museums; confirmed the open science engagement of major research organizations.

## Bibliodiversity

The diversity of those acting in scientific publishing. The main purpose of the Jussieu Call driven by the conviction that the open science policy should put an end to the dominance of a small number of scientific publishers imposing their terms to scientific communities.

## Boundary object

Describes a concept or an information that has a specific understanding in a local community of practice but is rigid enough to maintain its definition across communities too. Following Sam Moore, open access is such a boundary object: it has been successful because “*it resonates across communities but has specific meanings in individual circumstances, as a term multiple, processual and responsive to a range of motivations*”. Open science is probable another example of a boundary object.

## Color

An organizing feature of open access models.

- Green: open access via repositories; self-archiving or deposit by the author.
- Gold: open access via journal or book platforms. Different business models: e.g. APC, hybrid, Freemium (some services are chargeable, but access for free), sponsoring (learned society, research organization, foundation etc.).
- Bronze: articles with delayed open access on the journal platform, after an embargo period decided by the publishers and often with restricted reuse rights.
- Diamond: open access where neither readers nor authors have to pay; often with funding from academic institutions, learned societies, charities or government grants.
- Dark: Suber considers embargoed deposits in repositories as dark deposits; another meaning is open access by predatory publishers.
- Black: illegal open access that provides access to pay-walled articles.

## Community

A sociological concept to describe researchers' attitudes and behaviors. Built on shared practice, values and tools. Related and conditioned by disciplines, equipment, tools and projects; essential for ethics and learning. Data communities have been defined as "*fluid and informal network of researchers who share and use a certain type of data*".

## Data

Research data. Along with papers, significant part of the researchers' output. Subject to a life cycle. Requires curation. Stored in special repositories. Fuel of economy. Some researchers say that they don't produce data (but this needs further investigation). Probably the most important challenge for science. Defined by the 2019 EU Open Data Directive as "*documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results*".

## DMP

Data management plan. Description of what should be done with the collected and produced data. Dynamic, with different versions. Considered as good scientific practice. Mandatory for a growing number of funding bodies. Often including compliance with the FAIR principles. Tools like DMPonline are helpful for the writing of DMP. Could (should) be machine readable.

## Data paper

Authored, peer reviewed and citable article in academic or scholarly journals. A description of published research datasets, along with contextual information about the production and the acquisition of the data. Main purpose is to facilitate the findability, availability and reuse of research data. Part of the research data management and crosslinked to data repositories. For the moment, these papers represent less than 0.5% of all scientific articles.

## DOI

Digital Object Identifier. Sequence of characters that permanently and uniquely identifies a digital resource. Linked to a resource and its metadata. Allows a better sharing of the resource.

## DORA

San Francisco Declaration on Research Assessment or San Francisco Declaration. Statement initiated by scientists from the American Society for Cell Biology (ASCB) and a group of scientific journal publishers. DORA questions the increasing use of impact factor as an index

for the evaluation of research or researchers, and proposes the use of altmetrics. Published in 2013. Signed by more than 15,000 scientists and 1,550 scientific organizations (October 2019).

### Double dipping

Obtaining money from two sources at the same time. Describes a business model of publishers who get revenues via APC as well as via subscription fees, for the same products and from the same customers. Except for publishers, nobody appreciates double dipping.

### Embargo

A period of non-access to research results. The purpose is sometimes to protect authors' rights (e.g. PhD dissertations) and often to protect publishers' revenues (e.g. journal articles). Usually between 6 and 24 months. In some countries, the embargo period for publicly-funded research is fixed by law. During the embargo, the access is restricted to authorized users or to those who have paid for access.

### EOSC

European Open Science Cloud. EOSC is about connecting research infrastructures of data and publications. Yet, nobody can tell exactly what it is. The reason is that EOSC depends on the EU Member States' initiatives and contributions, supported by consortia of universities, research organizations, data producers and service providers. Work in progress.

### Ecosystem of open science

New. Emergent. Disruptive. A system with many stakeholders (publishers, libraries, research organizations, universities, funding bodies, governments...), with different functions (producing data and papers, dissemination of results, providing infrastructures or platforms, funding research projects and R&D...) and interests (financial, political...). However, in spite of all differences, the stakeholders are strongly interconnected and contribute in one way or other to the principles and goals of open science.

### ECR

Early Career Researcher. Includes generally PhD students, postdocs, sometimes also Master students. A paradoxical species: on the one hand, often considered as much in advance regarding digital tools and practice (digital natives); on the other hand, often not aware of research data management, data sharing and the challenge of open access to publications. An ERC requires special attention, needs empowerment and opportunities, and is often top of the priority of open science training programs.

### Elsevier

According to the Directory of Open Access Journals, the most important open access publisher, in terms of open access journals, followed by Sciendo (De Gruyter) and BioMed Central.

### FAIR

Findable. Accessible. Interoperable. Reusable. Guiding principles for good research data management. Good means above all, good for machines, i.e. good for the interconnection of infrastructures, platforms and so on. Identifiers, libre open access, standards, open licenses and rich metadata are essential for compliance with FAIR. The principles are widely approved and increasingly implemented for research data repositories. People start to apply FAIR principles also to the publishing of research papers.

### Funding body

Spends money for research projects conducted by others. Usually, the funding is conditioned by a call for projects and an expert selection based on criteria which can include elements of

open science, like data management, data sharing and open access to papers. Many people consider funding bodies as central stakeholders on the way to open science.

### Gratis

Gratis access is free of charge but restricted reuse rights. Indicated in Wikipedia with a green icon “free to read”.

### Grey literature

Unconventional or specialized documents (dissertations, reports working papers...), often unpublished, partly not peer reviewed. Useful for systematic reviews. A great potential for open access and bibliodiversity, with higher degrees of openness in institutional repositories. Largely neglected by open access monitoring.

### Harnad (Stevan)

Psychologist. Early “*archivangelist*” of open access and in particular of open repositories. Inventor of the open access colors green and gold. Editor of one of the very first open access journals (*Psycoloquy*). Main contributor to the first international declaration in favor of open access in 2002 at Budapest. Liked to write statements in form of numbered bullet lists. When angry or in despair, considered researchers as lazy and ignorant cowards.

### Hybrid

A successful mixed business model of journal publishing, combining subscription and APC. Readers pay for access while authors pay for dissemination. Risk of double dipping. Transitional. Not compliant with Plan S. Not about cars.

### Infrastructures

Fundamental facilities and systems serving research. Producing data, providing data services and other organizing structures and support for the institutions or communities they serve. Examples: particle accelerators, observatories, supercomputers, data repositories, journal platforms...

### Institutional repositories

Part of the green road to open access, through author self-archiving of their papers in institution (often university)-driven online collections or databases. A new vector of scientific information which have no counterpart in the traditional landscape of scholarly communication. The global directory OpenDOAR lists near to 4,000 institutional repositories in 2019. Not the only type of open repository but the most important one. Because of their potential for monitoring and assessment, convergence with current research information systems.

### Interoperability

Capacity of interconnection between machines, i.e. tools, platforms, infrastructures... Interoperability requires norms and standards, shared and common formats, terminologies, metadata, values etc. Interoperability is one of the main issues and purposes of the EOSC.

### Job profiles

Open science will change the job profiles of academic librarians, requiring new publishing, legal and training skills, and introducing the field of research data management. Should be appreciated as an opportunity of personal and professional growth but can be source of stress and worry. Needs change management.

### Journal (academic)

The rapid price increase of academic journals and the often slow and lengthy publishing procedures are two main drivers of the open access movement. Are journals part of the

problem? Open science strategies generally try to secure the 350-years-old model of academic journal publishing. Nobody wants the death of journals. But some people, like Egon Willighagen, start to think that perhaps, we should.

### Jussieu Call for Open Science and Bibliodiversity

An initiative of French information professionals and researchers to avoid the transformation of the actual scientific information market into a new open access oligopoly, involving the exclusive transfer of journal subscription budgets towards APC. The purpose is to develop and implement alternative models “*matching the aims of open science by asserting the need of supporting innovation for a thorough renewal of publishing functions*”. The call was adopted at Jussieu, the main campus of the Faculty of Science of Sorbonne University, Paris.

### Knowledge

The Berlin Declaration promotes the Internet as a functional instrument for a global scientific knowledge base and defines open access as a comprehensive source of human knowledge that has been approved by the scientific community. Open access requires the active commitment of each and every individual producer of scientific knowledge.

### Libre

Libre access is free of charge with large reuse rights (similar or near public domain). Indicated in Wikipedia with the orange icon “open access”. Often specified by Creative Commons licenses, in particular CC0 or CC-BY.

### Licensing

Following Wikipedia, a license is “*an official permission or permit to do, use, or own something*”. For open science, licensing is essential for two reasons: the agreements with publishers determine the pace and the nature of the transition to open access (transformative, disruptive...); and the conditions of the dissemination of open access papers and data via open licenses determine the reusability of research results (TDM...).

### Metadata

Data that characterizes and describes other data (digital objects) in a structured way (format). A major challenge for open science.

### Monitoring

Describes the development and implementation of policies, evaluates the effects, identifies potential gaps and outlines areas for improvement. The European Commission launched an Open Science Monitor to assess the development of open access to research results, by disciplines and countries, institutional policies, participative projects etc. Half of the EU Member States develop their own open science monitor, or will do so. According to the EU monitor, Slovakia ranks between Norway and Poland, with 46.7% open access publications, well above France, Germany and Italy (October 2019).

### Morrison (Heather)

Associate Professor at the University of Ottawa School of Information Studies and principal investigator on the Canadian SSHRC Insight Grant project *Sustaining the Knowledge Commons*. Her blog *The Imaginary Journal of Poetic Economics* is the first community-based monitor of open access to publications, with datasets on the development of open access journals, articles, preprints, metadata and other resources from 2005 on. The post recent update is from October 1, 2019. An essential source of knowledge on the open access movement.

## Network

Social network. A website that allows people to communicate and share information on the internet. A principal vector of open access to scientific papers, projects and data.

## *Nature*

*Nature* is a weekly international journal publishing “*the finest peer-reviewed research in all fields of science and technology on the basis of its originality, importance, interdisciplinary interest, timeliness, accessibility, elegance and surprising conclusions*”. Part of the *Nature Research* portfolio of 70 journals, all open access or hybrid, with APC ranging from €1,090 to €4,290. The *Nature* platform contains two open access data journals for data papers (APC: €1,180 and €1,390) and provides a Research Data Support Service for data sharing (€300 per dataset). A good showcase that open science does not mean the end of commercial journal publishing.

## Openness

Purpose as well as guiding principle of open science. Open means free accessibility; can be gratis or libre. In fact, different degrees of openness are observed, depending on reuse rights. Because of legal reasons, industrial interests, national security etc., politicians often add “*as often as possible (but) as closed as necessary*” when promoting open science and data sharing. Openness is nearly always considered as positive, desirable and right. One could imagine another approach to openness based on the digital experience that “*if it’s free, you’re the product*”; but who is the product of open science?

## ORCID

Open Researcher and Contributor Identifier. Persistent unique identifier. Distinguishes a researcher from every other researcher and, through integration in key research workflows supports automated linkages between a researcher and his professional activities.

## Plan S

Launched in 2019 by major funding bodies (cOAlition S), with help from academic publishers and the European Commission. The purpose is to make 100% open access mandatory in a short period. The guiding principles: dissemination of project results via 100% open access journals or open repositories, without any embargo. Hybrid journals are tolerated for a short transitional period. Plan S is considered as friendly for large publishers but not for learned society publishing. It has been estimated (a) that 85% of all academic journals are not compliant, and (b) that less than 10% articles are impacted. Dominique Babini from Argentina critiques the lack of global consultation and thinks that “*Plan S looks like an acceleration plan for Europe’s most privileged research institutions which can cover APCs for their research community, and ensure compliance with the Plan S requirements*”. Nobody seems to know what S really means.

## Quality

One major challenge of open science. How to bring together direct communication and quality control? Is peer review compliant with openness and transparency? Who will validate deposited datasets, and how? Will open science be quick and dirty? Some people think that powerful discovery tools, data analytics and artificial intelligence will be the solution.

## Replicability

Capacity of repeat an experience or observation. Refers to whether the results from a test or experiment can be replicated if repeated exactly the same way. Other researchers can test the findings of the research. The purpose is to confirm (or not) scientific hypotheses but also to detect unethical scientific practice, such as scientific fraud, falsified data, erroneous analysis

etc. Data sharing is generally considered as a way to avoid scientific fraud. Replicability does not (always) require the original data. Reproducibility does (always).

### Researcher

Major element of the ecosystem of open science. Producer of data and papers. Object of evaluation. Should be working on the campus but may be somewhere else. In the past has been seen in ivory towers. Member of communities. Not always aware of the challenge of open science, sometimes described as a potential threat to open access. Need more attention and further investigation.

### Reusability

The possibility of reusing data produced by others. Reuse can mean new analyses, a new interpretation, merging with other datasets etc. Accessibility is not enough if you do not know the meaning. Requires rich information (metadata) about data collection and production, about their meaning, and about the conditions of use (licensing). Data papers can be helpful to improve reusability.

### Science

For some people, science is the final purpose of open science; open science is a way to improve the work environment of researchers. For other people, science is a complex and costly organization which requires more efficiency, more transparency, more inclusiveness, higher speed and productivity and less spending; open science is a way to get there and make researchers do more with less. Finally, some people consider science as a producer of useful data, information and knowledge for society as a whole and industry in particular; open science is a way to provide them with new business opportunities and gratis (and if possible libre) fuel.

### Suber (Peter)

Philosopher, director of the Harvard Office for Scholarly Communication and of the Harvard Open Access Project. One of the best-known spin-doctors and leading voices of open science and especially of open access to research papers. His conviction: *“The basic idea of open access is simple: Make research literature available online without price barriers and without most permission barriers”*. His reference work on open access is freely available on the MIT Press website, since 2012.

### Sustainability

One of the criteria of open science strategies. Sustainable means that funding bodies will provide money for the development and launch of new tools and services but not for further, operational costs. Sustainable means also that corporate companies want to secure or increase their revenues. The period of sustainability is variable; some speak of 2-3 years, others of 5-10 years.

### TDM

Text and data mining. Has been defined as *“the discovery by computer of new, previously unknown information, by automatically extracting and relating information from different resources, to reveal otherwise hidden meanings”*. A vital requirement in a world of big data and information overflow. In the context of open science, TDM is a main challenge of legislation, licensing and business models, especially as a specific reuse of freely available research data and papers.

### Transformative

Agreements with publishers are called transformative if they contribute to attain 100% open access publishing in a short time period. Two major options for transformative big deals: “read-



and-publish” (settle the global price of the subscription to a collection of closed or hybrid journals; the researchers can publish in open access without or reduced APC) and “publish-and-read” (settle the global amount of APC for a collection of open access or hybrid journals; the researchers can access all journals, open access or not).

## Twenty twenty

The target of the Amsterdam Call was 100% open access in 2020, with regular monitoring and evaluation. 2020 was also initial deadline of Plan S for 100% open access to research papers. 2020 has been acclaimed as “action-provoking” and disruptive but has also been considered as unrealistic. 2020 is now. The goal of 100% open access is not attained, and new deadlines are fixed.

## Unpaywall

Open database run by Impactstory, a nonprofit organization. Provides information about open access status of more than 24.5 million articles (October 2019). Harvests content directly from over 50,000 publishers and repositories. Useful for open access monitoring.

## Values

Openness, i.e. the personal and professional commitment in favor of open access to information and knowledge, is part of the fundamental values of information professionals. Clause 4 of the IFLA Code of Ethics says that “*librarians and other information workers' interest is to provide the best possible access for library users to information and ideas in any media or format. This includes support for the principles of open access, open source, and open licenses*”.

## Worldwide

Open science is usually considered as universal, as a global way to do better science. However, some people think that open science may confirm the dominant position of Western science. In the past, researchers of the Global South had problems to access scientific information, because of subscriptions; in the future, they may have problems to publish their results, because of APC. Some people think, too, that researchers from the Global South should not give away their results for free, and that they should retain control about their data, information and knowledge.

## arXiv

The mother of all open repositories; the one who rules them all. Launched by the global high energy physics community in 1991, initially hosted by Los Alamos, now at Cornell. Its central purpose is to enable direct communication of research results between researchers. Requires the deposit of preprints (or e-prints); no records without the full paper. In October 2019, arXiv contains more than 1.6 million papers, mainly in physics, mathematics and computer science.

## You

I.e. knowledge producer, information professional, academic librarian, data officer, archivist or other: you are partners and support of researchers, and essential stakeholders in the ecosystem, for the promotion and achievement of open science.

## Zenodo

Open repository from CERN. Contains in October 2019 more than 1.4 million items, mostly publications but also datasets, image files and software, from CERN and nearly 3,700 other scientific communities. Zenodo is a model of a public, sustainable, open software (Invenio), multidisciplinary and multi-resource research repository. Zenodotus was the first librarian of the Library of Alexandria.

## Zika

A dangerous viral disease. Along with Ebola and climate change, one of the main arguments of politicians during the preparation of the Amsterdam Call in 2016. The idea is that open science will help to discover better and faster solutions for crucial societal problems, especially by providing access to the most recent scientific knowledge for a broad group of potential contributors, including new or unknown users of knowledge.

### Contact address

Dr. Joachim Schöpfel

University of Lille

GERiiCO Laboratory

BP 60149

59653 Villeneuve d'Ascq Cedex

France

[joachim.schopfel@univ-lille.fr](mailto:joachim.schopfel@univ-lille.fr)