Evaluation and Open Science in Social Sciences and Humanities

Nabil Hathout

CLLE, CNRS & Université Toulouse Jean Jaurès

Masterclass Science Ouverte
Grenoble, June 4, 2024

Nabil Hathout

- Directeur de recherche CNRS
- Linguist, morphology and natural language processing (NLP)
- **Previous** chair of the linguistics section of the Comité National de la Recherche Scientifique (CoNRS, section 34, 2016–2021)
- Currently member of the research commission of Université Toulouse Jean-Jaurès (2023–)

This presentation benefited from discussions with :

- Mehdi Senoussi (CLLE) ;
- the attendees of the CLLE workshop "Doctorants-Permanents" organized by Hélène Cochet (CLLE).

Evaluation in social sciences and humanities

Open science in Evaluation

Evaluation in an Open Science framework

Thoughts on evaluation and open science

Evaluation in scientific research

Evaluation plays an important role in science. It guides most of the decisions regarding the publication of results, hiring, funding, PhD defense, etc.

Several types of evaluation:

- PhD thesis $\Rightarrow$ Evaluated by 2 or 3 *rappoteurs* and a jury.
- Articles submitted to journals and conferences $\Rightarrow$ Peer reviewed.
- Projects, funding application, grants (ANR, ERC, Horizon Europe) $\Rightarrow$ Evaluated by experts and committees.
- Activities of candidates that apply for a job or a promotion (university, CNRS, INSERM, etc.). Periodic evaluation of researchers (CNRS) $\Rightarrow$ Evaluated by committees.
- Teams and labs (HCERES) $\Rightarrow$ Experts.

Evaluation in scientific research

Evaluation is always **commissioned** by an institution (job application; funding) or a board (journal, conference).

Evaluation is relative to the **criteria** of the commissioning institution or board.

Evaluation depends on the **quality of the experts**.

- Not enough senior experts to evaluate all the scientific output in disciplines where the number of submissions increases too rapidly (e.g., artificial intelligence).

Reminder

If we assume that a submission is reviewed **3 times**, for each article submitted, a researcher should review **3 papers** for the peer review model to work smoothly.

Evaluation of individuals is difficult

Katalin Karikó, Nobel Price winner (Physiology or Medicine; 2023) was not promoted full professor at UPenn because her funding applications were not successful enough.

Evaluation must take into account of a wide range of activities:

- publication in selective journals
- success in funding application
- team creation and lead
- doctoral supervision, teaching
- contribution to collective interest tasks

In France, committees must conform to the legal framework for civil service competitions.

Evaluation of individuals is difficult

Committees must compare (and often rank) the work of candidates that conduct research:

- in different sub-disciplines, and
- of different kinds (theoretical, descriptive, experimental, applied)

How to compare the activities

- of a linguist working in theoretical syntax,
- a field linguist collecting data on languages for 6 months a year, and
- a psycholinguist using complex equipments such as fMRI?

Evaluation of individuals is difficult

An easy solution is to order activities in the same way for all disciplines.

In general, **publication** is considered the most important activity despite the fact that **the primary purpose of research is the creation of knowledge**, not the production of publications.

Side effect. Evaluation criteria determine researchers' **behavior**. Researchers prioritize publication in selective journals and funding because these are the most profitable outputs.

Evaluation in social sciences and humanities

Bibliometric criteria are less important in SSH than in other sciences:

- A large part of SSH publications are not indexed in Web of Science and have no journal impact factor. They are not published in Springer or Elsevier journals.

Methods and practices may vary a lot from one discipline to the other (neuropsychology *vs* choreographic art).

In a number of SSH disciplines, most publications are in French.

Publication in French, Spanish, Italian, Portuguese, Russian, Mandarin, etc. are considered to have the same importance as the publication in English.

Evaluation in social sciences and humanities

Open science in Evaluation

Evaluation in an Open Science framework

Thoughts on evaluation and open science

Open Science recommendations

Committees that evaluate individuals tend to delegate the evaluation of the content of the researchers' work to journals and conferences.

DORA (San Francisco Declaration on Research Assessment)

Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

DORA is signed by many institutions, including CNRS and UGA.

Open Science recommendations

... BUT

evaluation by CoNRS sections (and many other committees) is carried out by scientists who are not familiar with open science practices.

CoNRS sections and juries do as they see fit. They generally just follow the traditions and practices of their field and sub-fields.

The importance attached to the prestige of journals remains the rule.

This can be seen in remarks like **"there are only two international journals"** in section reports, meaning that there are only two articles in Springer or Elsevier journals.

- Predatory journals (e.g., MDPI, Frontiers) are often better evaluated than good French journals.

There is a lot of room for improvement.

Open Science recommendations

DORA recommendations are difficult to follow:

- Committees are unable to evaluate the content of job or funding applications because of the number of candidates, the limited expertise of the members, etc.
- Experts rely on the evaluation carried out by the journals and the conferences.
  - Their reviewers are more specialized.
    They know the expectations and practices of their field.
  - Acceptance assesses the quality of the paper.
    - Committees do not have access to the review reports.
  - Implicitly, work published in more selective journals is considered to be of higher quality.

Open Science recommendations

Bibliometrics give an impression of objectivity

Institutions are in competition with each other at all levels, and can put forward bibliometric indices as indicators of the success of their researchers. A more "open science" evaluation would make comparison more difficult.

At the CNRS, some assessments (e.g., for promotions) are based on **5 or 10 articles** chosen by the candidates, so that the files do not contain a complete list of the candidate's scientific output ...
but the forms still include a field intended for the URL of a document that provides a complete list of the candidate's scientific output, or a full CV.

For the same reasons, the evaluation period is also limited (usually to the last 5 years).

Open Access

The first open science recommendation is to make articles and research output available on repositories such as ArXiv or HAL, on researchers' personal pages, on the sites of the journals, etc.

CNRS researchers are required to deposit all their publications on HAL.

Compliance with OA recommendations is not self-evident in some SSH disciplines

- Many results are published as books by relatively small, highly **specialized "niche" publishers**.
  - Authors generally have to **transfer their copyright** to the publishers, and often have to pay additional fees.
  - OA may jeopardize the viability of these publishers.
- Works is designed for publication **in print**.
- Work may include **iconography**, which imposes constraints on printing.

Open Access

Open access may be difficult to implement for work that falls within the scope of

- **creation research**, which may take the form of a recorded artistic performance, or
- **action research**, e.g., fieldwork with local players with the aim of creating a cultural animation in a community.

Performance (resp. animation) is both the product of research and its object of study.

Open Science in CoNRS section criteria

Open science practices are explicitly mentioned in the 3rd item in the current criteria of **section 34**.

*Qualité des activités de dissémination auprès de la communauté scientifique (**mise à disposition des données ou de logiciels**, participation à des conférences reconnues, congrès ou séminaires...)*

with a note:

***Type de mise à disposition (archives ouvertes, banques de données, etc.)**, type d'interventions (communications orales, affiches), capacité à impliquer des membres de son entourage scientifique, importance du congrès dans le domaine de compétence pour les interventions les plus significatives, etc.*

But open access is not mentioned.

Open Science in CoNRS section criteria

Open science practices are not mentioned in the criteria of the archaeology and history sections (32 and 33).

They are not mentioned in those of section 35 (philosophy, literature, art).

They are absent from the criteria of more experimental science sections such as psychology (26), sociology (36) and economics (37).

Open Science in CoNRS section criteria

Open science best practice

is not currently an important criterion in the evaluation of individuals and teams.

Open access publication, open data, open source code may be required by funding agencies.

Open Data

Making research data available is often difficult. It can be requested by journals and conferences in experimental disciplines such as cognitive science, psychology.

Funding agencies (ANR, ERC, etc.) usually

- require that all the articles produced as part of projects they fund be disseminated in open access, and
- ask for **data management plans** (DMPs) that include the release of data and its long-term preservation.

Complying with open science recommendations represent a costly overhead for the researcher.

Their cost-effectiveness should be improved by making them mandatory for all research.

How to take this overhead into account ?

Open Data in SSH

In the social sciences and humanities, research is often individual or carried out in small groups or groups of variable geometry within the funded projects.

This form of research makes the burden of preparing and disseminating research data very heavy.

It is easier when carried out in teams with support staff (especially engineers) who can help with these tasks.

Open Data in SSH

Much SSH data is copyrighted or difficult to access, for example when held in archives.

Many data contain information about individuals and need to be **anonymized** before they can be released (RGPD).

Many data have been collected without any thought that they might one day be disseminated openly, and for which the authors have not asked for and obtained the **informed consent** of the people interviewed, recorded, etc.

Corpora and projects are built up over relatively long periods. Data may be used 10 or 20 years after their creation.

$\Rightarrow$ It is not conceivable to include open science practices as criteria for the evaluation of such work.

Open Data in SSH

Open science practices help to limit **misconduct**, such as that of psychologists Dan Ariely and Francesca Gino, highlighted by Data Colada.

Open science practices are essential for the replication of studies. Replication is not relevant in some disciplines, especially in scholarly work, action research, creative research, etc.

Open science practices are essentially formal criteria that do not guarantee the quality of research.

We do not know whether, in SSH, knowing that research data will be released makes them more carefully produced, and whether this leads to more reliable results.

An evaluation of open science practices on research quality has yet to be carried out.

Evaluation in social sciences and humanities

Open science in Evaluation

Evaluation in an Open Science framework

Thoughts on evaluation and open science

Criteria

Many scientific activities fall outside the scope of open science:

- Funding applications
- Teaching
- Organization of conferences
- Mobility

Several journals and conferences provide detailed guidelines for reviewing submissions. The criteria used by funding agencies like ANR or FWO (Flanders) are very precise.

The criteria of the CoNRS sections are more general in order to limit the juridic risks. The quality of feedback and the informativeness of the reports may be limited for the same reasons.

Criteria depends on the discipline, on the country and its tradition. In linguistics, in France, few or no reviewing criteria are given by renown journals like *Language* or *Langue Française*.

Code and Data

We do not know how to assess the value of code made available on git repositories.

Most researchers and most experts in SSH are not trained in software engineering.

Assessing the form and impact of data is a complex task.

The **usefulness** of data, like that of publications, can only truly be estimated over the medium term.

This period does not match that of the evaluation of individuals based on their most recent productions (typically those of the last 5 years).

Evaluation in social sciences and humanities

Open science in Evaluation

Evaluation in an Open Science framework

Thoughts on evaluation and open science

Open Reviews

Review reports should become available for all and be **attached** to the articles.

Separate the functions of journals and conferences

- **rating**
- **dissemination**
- **social**

Rating should be made public for all the submissions that are evaluated.

Evaluations will only be carried out once.

Dissemination is not a problem anymore.

Workshops and conferences would become smaller (with lighter footprints), more specialized, more focused on the creation of communities and networks.

Other thoughts

Training

Teach **collaborative methods** to Master and PhD students. These method could be adapted from open source development practices.

Result-blind peer review

Promote the publication of research work in the form of **registered reports**.

- Journals accept or reject papers before the results or conclusions of the study have been made.
- Submission describes the theoretical justification, experimental design, and statistical analysis of the study.

Other thoughts

Evaluation of individuals

Evaluate the content of researchers' work by calling in **external experts**. This proposal is not compatible with the way researchers are currently hired and evaluated in France.

Research more, publish less

Limit the number of papers a researcher can submit each year.
It is impossible to evaluate the work of researchers that publish up to two or three papers a week.