



HAL
open science

Study on the quantitative assessment of performance in Social Sciences and Humanities

Christine Kosmopoulos

► **To cite this version:**

Christine Kosmopoulos. Study on the quantitative assessment of performance in Social Sciences and Humanities. [Research Report] Final report assigned by the ERC Expert Group for Programme Monitoring and Evaluation, Brussels, ERC. 2015, 45 p. hal-01853406

HAL Id: hal-01853406

<https://hal.science/hal-01853406>

Submitted on 16 Jan 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.

**Study on the quantitative assessment of performance
in Social Sciences and Humanities**

Final Report of Christine Kosmopoulos for the ERC

CNRS-UMR 8504, Paris, France

Assigned by the ERCEA

For

The ERC Scientific Council Working Group

For Key Performance Indicators



Final Report, June 2015

- Copyright ©European Research Council, 2019
- Reuse is authorised under Creative Commons Attribution 4.0 Licence (<https://creativecommons.org/licenses/by/4.0/>)
- This study/project has received funding from the European Research Council (ERC) under the European Union's H2020 research and innovation programme.
- The analyses and findings of this article/paper are under the responsibility of the Authors and do not necessarily reflect the views of the European Commission, ERC Executive Agency or the ERC Scientific Council, which cannot be considered responsible for any errors or omissions.



Table of Contents

The objectives of the study.....	4
1. Quantitative Assessment - A short historical recall.....	6
1.1 ISI Citation Indexes.....	6
1.2 The transition to the quantitative evaluation.....	8
2. The present sources of bibliometric data for the SH.....	10
2.1 The coverage of <i>WOS-AHCI</i> and <i>WOS-SSCI</i>	10
Conclusion for the <i>WOS (ACHI and SSCI)</i>	11
2.2 The coverage of <i>Scopus</i>	11
Conclusion for <i>Scopus</i>	11
2.3 The coverage of <i>ERIH</i>	12
Conclusion for the <i>ERIH</i> lists.....	13
2.4 <i>Google Scholar (GS)</i>	13
2.5 <i>Publish or Perish (PoP)</i>	14
Conclusion for <i>GS</i> and <i>PoP</i>	15
2.6 Evaluation activities of national research systems: The Case of France.....	15
Conclusion for the <i>AERES'</i> lists.....	16
3. The accuracy of the bibliometric data for the SH.....	17
3.1 The <i>JournalBase</i> project.....	17
3.2 A comparative study of the <i>WOS, Scopus, ERIH, AERES</i>	18
Conclusion of section 3.....	22
4. What are the specificities of the SH?.....	24

4.1 The types of production according to the different fields of the SH.....	24
4.2 The pre-eminence of presentations at congresses	30
4.3 The linguistic diversity	30
Conclusion of section 4	32
5. New practices: Science in Open Access	34
5.1 The emergence of Open Access (OA) practices	34
5.2 The different models of OA’s scientific content	34
5.3 The services developed with OA	37
5.4 The “Open Sources”	37
5.5 The interoperability of the data	37
5.6 Altmetrics	39
Conclusion of section 5	44
6. Relation between quantitative and qualitative evaluation in fields of Social Sciences and Humanities	45
7. General conclusion	48
8. Recommendations	50
8.1 A database dedicated to the SH.....	50
8.2 An Open Access Repository	51
8.3 The final aim: a big data repository associated to the peer-review	52
8.4 The advantages of such a tool for the ERC	53
Bibliography.....	54
Annexes	56
Annex 1: Description of the current metrics of <i>Publish or Perish</i>	56
Annex 2: 27 headlines of <i>JournalBase</i>	58

Annex 3: The number of journals per bibliometric sources and per discipline	59
Annex 4: The 49 languages retrieved from <i>JournalBase</i>	61
Annex 5: Number of journals involved in the use of the indicated language	63
Annex 6: A sample of <i>Zotero</i> 's network.....	66
Annex 7: A cloud of key-words generated by publications	67

The objectives of the study

As part of the ERC Monitoring and Evaluation Strategy, ERCEA is conducting various analyses of the results (mainly publications) produced by the ERC projects. These often involve quantitative assessments of results, e.g. bibliometric and altmetric measurements. The monitoring and evaluation of the ERC projects in the Social Sciences and Humanities (SH) domain pose specific problems. This is not only a matter of defining bibliometric indicators for SH domain, but also of access to publications as both Web-of-Science and Scopus include few journals in SH domain, mainly English ones, and a limited number of books.

I have been assigned to conduct small scale analysis and studies on selected topics based mainly on the ERC's Monitoring and Evaluation strategy, but also to take into account good practice and experience from monitoring and evaluation activities of national research systems, in particular regarding the ERC's complementarity and added value of the ERC Expert Group for Programme Monitoring and Evaluation.

The purpose is to identify and present the current practices related to the quantitative assessment of performance in Social Sciences and Humanities, especially regarding:

- sources of bibliometric data in the SH domain making a distinction between SH fields when necessary,
- choice of performance indicators by type of publication (e.g. article, monograph, book / book chapter) and by SH field,
- altmetrics or other quantitative indicators of performance,
- relation between quantitative and qualitative evaluation in fields of Social Sciences and Humanities.

I, as well, have been asked to focus my study on those practices that could also be applied by ERC and conclude it with recommendations to ERC.

In addition, during the meeting of 28th January 2015, I have been assigned:

- to study the different types of publication including the unorthodox production
- to study the question of Open Access in order to track down to what degree the practices have changed

1. Quantitative Assessment - A short historical recall

In this Section, I shortly recall this history of the quantitative assessment in Science, in order to understand how it was developed, what it has achieved and the different aspects of its use and influence in scientific research.

1.1 ISI Citation Indexes

In the 1960's, Eugene Garfield -founder of the *Institute for Scientific Information* (ISI) launched the famous database *Science Citation Index* (SCI)¹ with the support of the *National Institute of Health* (NIH) and the *National Science Foundation* (NSF), in the United States. The *Science Citation Index* was based on the *Current Contents*² already developed in the 1950's. They got their final name of *Current Contents* in 1958.

Originally, the *Science Citation Index* was not supposed to evaluate scientific production, even if the citation data made available for the first time a quantitative assessment. At the beginning, it was a documentary instrument (mainly devoted to Life Sciences and greatly used in Medicine). Here it was a question of making known the network of scientific references and facilitating access to the information on a given theme in different disciplines, with the advantage for the sciences of having simultaneous access to the work of both applied research and fundamental research. Through the citations produced in articles, it was possible to find articles dealing with the same subject, whereas librarians' traditional indexing system provided entries which were too general.

As has been mentioned in several articles of E. Garfield, the purposes of the *Citation Indexes* were to bring a new dimension in documentation through association of ideas (he recalled it in

¹ According to the British indexing standard (BS3700:1988), an *index* is a systematic arrangement of entries designed to enable users to locate information in a document. The process of creating an *index* is called *indexing*, and a person who does it is called an *indexer*.

² They reported on a certain number of what we would call today "Metadata" such as the title of the article, the authors, the abstract, the bibliographical information on which the article was based, the name of the Journal, its volume, number and so on. The CC were the unique source of information for the scientists all over the world, to know what is being studied and what has been published in their particular domain. The CC were mostly consulted in natural sciences. The subscription to the *Current Contents* was and is still very expensive.

an article in 2006), to close the gap between fundamental and applied science (Garfield, 1963a), to cross artificial boundaries imposed by classification schemes and journal titles (Garfield, 1963b)³ and to encourage the interdisciplinarity in order to extend the limits of the scientific fields. The objective as announced was highly pertinent: to make an inventory of the sources cited in an article, as well as the references correlated after its publication.

It is undeniable that the achievement of this expert documentation, the *Current Contents* and the *Science Citation Index*, was of great interest. Most scientists could not work without them, even today. Following the movement, the scientists started to base their own publications on the sources provided by the *Current Contents* and to publish their articles in the list of Journals indexed by *ISI* in order to get a greater impact for their publications, and to appear as high as possible in the *Science Citation Index*... It also then helped to build an 'official' list of scholarly Journals in the fields included in *Science Citation Index*, a list selected by *ISI* under its own criteria.

Thereafter *ISI* developed the *Social Sciences Citation Index* (SSCI) and later the *Arts and Humanities Citation Index* (AHCI). The databases of the *Citation Index* allowed analyzing the impact of the Journals listed by *ISI* and thus the idea to study the impact of the Journals naturally came up, hence, the *Journal Citation Report*. It became rapidly a part of the *Citation Index* and since 1979 the *Journal Citation Report* explores the so-called *Impact Factor* of the Journals, extracted from the *Science Citation Index*.

For the record, I definitely would like to quote E. Garfield about the tools he developed with the sponsors of the *NSF* and the *NIH*: "It should be emphasized that the basic purpose of the project is not to take a statistical inventory of scientific production. That is, indeed, an important byproduct of the work. The main objective, however, is to develop an information system which is economical and which contributes significantly to the process of information discovery – that is, the correlation of scientific observations not obvious to the searcher. Citation Indexes can provide new insights impossible through descriptor-oriented systems"(Garfield, 1963c). However, as a result of the sponsoring of the *NSF*, this base had the stated mission of assisting those making assessments to judge the pertinence of the bibliographical sources of the articles submitted to the journals (Garfield, 1964).

³ For example, consider Einstein's 1905 paper analyzing Brownian movement and diffusion (Ann. Physik. 17, 549). This paper was cited in 1961 not only in chemistry, mathematics and physics journals but also in a biological study on the immunological application of double diffusion (C.J. Van Oss, Zeit. Immunitätsforsch. 122, 44 (1961))

1.2 The transition to the quantitative evaluation

In 1980 the *NSF* published a first report with bibliometric indicators based on the *SCI* in order to establish a scientific policy. From that period, the *Science Citation Index* and the *Impact Factor of the Journals* started to play a major role in the scientific assessment in some of the so-called ‘hard’ and natural sciences. In medical research, for instance, they became totally unavoidable. Since 1980, the national and the regional indicators use the *SCI* data “to identify trends of one kind or another” (Garfield, 1998).

In 1992, the private company of E. Garfield, *ISI*, was acquired by *Thomson Reuters*. The *Thomson Group* developed on the web the well-known *Web of Science* which includes the three different citations databases: *SCI*, *SSCI*, *AHCI*, to which are associated a series of bibliometric indicators⁴. These quantitative tools were intended to measure the quality of the work produced by the researchers. Based on the *SCI*, they are now implemented in most of hard and natural sciences and frequently used by the Institutions, Universities and Funders etc. all over the world to evaluate the “objective impact” of the research outputs.

Parallel to the traditional Peer Review, quantitative tools are intended to measure objectively the performance of the work of countries, institutions, researchers etc. by allowing comparisons using measurable indicators. For more details on the development of algorithms, data sets, and bibliometric tools by *ISI* and later *Thomson Reuters*, please refer to the article of D. Pontille and D. Trony (Pontille, Torny, 2013).

However, one would have to check the accuracy of the data used to produce the measurable indicators (section 2 of this report). As a matter of fact, at the origin of any database construction is a human brain. The purpose of the data and their nature are always designed by human intelligence including necessarily a part of its own subjectivity. In case of the *Citation Indexes*, they were not built for a bibliometric use and the initial documentary goal has been devious. This derivation causes many difficulties. For the physicist Frank Laloë to use the *Citation Indexes* as basic elements for the quantitative evaluation contradicts their real function. Even worse, they might change the practices of the citation’s costs to the qualitative assessment (Laloë, Mosseri, 2009). As we have seen (section 1.1) their real function is documentary and based only on the journal's articles. It makes it very restrictive and does not reflect the whole production that should be included in any quantitative assessment.

In 2009, even E. Garfield warns us “The new generation of scientists and even scientometricians need to be regularly reminded that the *Science Citation Index* (*SCI*) was not

⁴ Bibliometry is the application of mathematics and statistical methods to books, articles and other means of communication.

originally created either to conduct quantitative studies, calculate impact factors, nor to facilitate the study of history of science.....The growth of COLLNET and ISSI and this symposium demonstrate that it is appropriate to say of citation analysis, that not only is the tail wagging the dog, but that the tail has become a huge animal which is rapidly mutating into a “multiplegic monster.” “(Garfield, 2009). Here scientometrics have to be understood as the Tail and the information retrieval (the first aim of the citation index) as the Dog...

Considering what is happening today, it is undeniable that in natural and medical sciences, scientometrics became an inevitable part of the evaluation for about 35 years. We entered then a new paradigm in the meaning of the scientific research, in terms of definition, goals, performance, analysis and assessment. This new paradigm is being reinforced by the functionalities of the Internet.

2. The present sources of bibliometric data for the SH

In this Section, I make as exhaustive as possible an inventory of the different sources, based on several publications. I also study an example of national research system for the assessment in SH and how they use or not these sources.

2.1 The coverage of *WOS-AHCI* and *WOS-SSCI*

Although considered by some to be the standard citation database, the *Web of Science* of Thomson Reuters has been criticized for many years for its bias against the SH. A report by Philippe Jeannin (Jeannin, 2003) on *the quantitative evaluation of research in the SH* submitted in 2003 to the French Ministry of Research and New Technologies confirms that French journals in these domains were not covered. In 2009 the *WOS* inventories some 8,700 International journals, but only 3,000 for the SH, almost all of which are of Anglo-Saxon origin (Kosmopoulos, Pumain, 2007).

Four years later, in 2012, the proportion did not change. The *Art and Humanities Citation Index* (AHCI) covered 1,894 journals and the *Social Sciences Citation Index* (SSCI) 2,838 journals gathering in total 4,097 journals (Kosmopoulos, Dassa, 2013) for SH on a total of 12,500 journals indexed by the *WOS* taking into account the *Science Citation Index* (SCI). Whatsmore, if we look closer, some journals are both included in AHCI and SSCI which means that in fact there are even less. Actually if we adjust the total to unique titles of journals, there are between 3,500 and 4,000 journals altogether for SH in the *WOS*, although the total amount of Peer-reviewed Journals in SH referenced in *JournalBase* exceeds the 13,000 titles⁵.

It is interesting to note that none of the SH journals supported by the *Centre National de la Recherche Scientifique*⁶ (CNRS, France) appear in the *WOS* until today (March 2015), although the CNRS tried several times to negotiate improvements with *Thomson Reuters/ISI* that would take into account specific European needs.

⁵ Journalbase.cnrs.fr. For more details, see section 3.1 of this report.

⁶ journalbase.cnrs.fr

Conclusion for the *WOS* (*ACHI* and *SSCI*)

At this stage it has been totally proved that the *WOS* covers only a very few number of the peer-reviewed journals in SH and that it cannot be used alone for evaluating the impact of an article in SH Journal. As history teaches us (section 1.1), the *Citation Indexes* were mainly created for the natural and medical sciences and not the others. That might partially explains why the sources remain so incomplete.

If we consider the group “United States and United Kingdom”, it is confirmed that the commercial databases of the *WOS* inventory a large majority of Anglophone journals, in any case more than 2/3 of the journals, and up to almost 80% for the *WOS-SSCI* database (Kosmopoulos, Dassa, 2013).

2.2 The coverage of *Scopus*

The Dutch publishing group *Elsevier* shook the world monopoly held until then by *Thomson Scientific*, when they launched the citations database *Scopus* (2004). The ambition of this commercial publisher was to open the way for access to the whole of world institutional scientific literature. Actually, the number of journals posted on the *Scopus* website in 2008 was impressive: for the SH compared to the *WOS ACHI* and *SSCI*, it was about 6,200 titles (the initial general file of *Scopus* inventories 22,688 journals, of which 15,863 were active for the whole of the sciences). In 2012, they claim to include 6,625 Journals for the SH (Kosmopoulos, Dassa, 2013).

Conclusion for *Scopus*

Even if *Scopus* covers 6,625 Journals in 2012, that is to say around 2,000 more than the *WOS*, its coverage remains about half of the total number of journals indexed in *JournalBase* (over 13,000 – about *JournalBase*, more information is provided in section 3).

Furthermore, the bibliometric tool *Scopus* has been developed by Elsevier in a very short period and is not based on rigorous indexing. A lot of mistakes have been underlined (Dassa, Kosmopoulos, Pumain, 2010). Since 2010, all the initial key-words up to 199 in 2008 have been removed and no precise disciplinary category was settled on in order to classify the journals (Dassa M., Kosmopoulos C., 2010, note 11).

To access the data of *Scopus* and of the *WOS*, there is a high fee which is negotiated (at the full price) at the level of the management of the research organizations, university libraries, or

founders. For instance in 2015, France concluded an agreement of 17 million euros for access to the Journals of *Elsevier* and partly to *Scopus*. It is therefore important from the financial perspective, as well, to know their exact coverage for the SH and reliability.

2.3 The coverage of *ERIH*

In 2007, in reaction to the hegemony of the *WOS* and in the absence of enough representation of databases for non-Anglophone SH journals, the *European Science Foundation (ESF)*, in collaboration with the consortium *HERA*⁷, for its part, threw itself into making inventories of the European journals in SH, with the help of panels of experts, and published lists of journals, and again in 2008, according to the fields of discipline (*ERIH*)⁸ for the Humanities⁹. In 2012, the total number of the journals indexed in *ERIH* was 5,317 entirely dedicated to the Humanities. But many other disciplines of social sciences, such as Geography, Economics... are not concerned by the *ERIH* lists.

In 2014 the *ESF* concluded an agreement with Norwegian Social Science Data Services (NSD) and developed *ERIH Plus* including for the first time the Social Sciences' journals. *ERIH Plus* is still under development. In May 2015, the list of the disciplines and journals is not yet finalized¹⁰. For this reason, my remarks are based on *ERIH*' lists (from 2007 till 2014) and only concern the Humanities.

⁷ ESF: the European Science Foundation, founded in 1974, is a non-governmental association which gathers 80 member organizations dedicated to scientific research in 30 European countries. HERA: a consortium of funding agencies for research in social sciences.

⁸ <http://www.esf.org/index.php?id=4813>.

⁹ The definition of Humanities has not yet reached a consensual agreement within the international scientific community. In some countries, there are disciplines included in the Humanities, and not in others which consider them more as Social Sciences...

¹⁰ For example, Geography is not part of the disciplinary list at that date.

Conclusion for the *ERIH* lists

The total amount of Journals in *ERIH* lists reaches 5,317 for the Humanities, while at the same time, the *WOS/AHCI* indexed only 1,894 journals. At this stage *ERIH* is the best resource that can be used to check for high quality scholarly journals in the Humanities.

However *ERIH* only provides lists of selected scholarly journals but no bibliometric tools. The evaluation A, B, C, applied to each journal at the beginning, in order to classify the impact of the journals, has been quickly removed. Actually one can check whether a journal is included or not in *ERIH* lists, but no statistical indicators of impact are available.

The *ERIH* lists are public and in Open Access.

2.4 *Google Scholar (GS)*

Another source launched at the end of 2004, *Google Scholar (GS)*, is not strictly speaking a bibliographical database in its initial conception; it is a search engine specializing in scientific literature. The database that grew out of it is the property of the group that launched the Google search engine in 1998. The indexed documents originate with scientific editors, scholarly societies, collections from open archives, universities and other research organizations. *Google Scholar*, like Google, uses a robot tool that proceeds to explore, classify and index the contents of a proposed site.

A real advantage of *Google Scholar* for the SH, in addition to the free access, is that it inventories all scientific literature without distinction: articles in journals, whether or not they have an editorial board, but also theses, books, extracts of books, reports, pre-prints, etc. This is one of the reasons that it is usually consulted by various Institutions and Researchers, especially in the SH domains. It is the only tool where someone can easily get a rough idea of what someone is producing and how many times a reference is cited.

Nevertheless, this specialized search engine presents other problems. Unlike the *WOS* or *Scopus*, it provides no information about the resources it uses and no detailed lists. A study that I did a few years ago –still valuable- (Kosmopoulos, 2009), shows that there is no list of the commercial editors or of the servers of the indexed archives, and no information about the period covered, the volume, or even the countries involved. Certain known sources of *Google Scholar*, for example the bibliographical database Francis of the INIST (France), are not in conformity with bibliometric norms and some of their references are not mentioned at all. Actually, there are also significant gaps in the coverage of publishers' archives. A test with reference to important figures in the sciences confirms this observation, as Peter Jacso demonstrates (Jacso, 2005).

2.5 Publish or Perish (PoP)

In 2007, Ann Will Harzing, an Australian academic specializing in information technology, developed independently a tool for text treatment and citation analysis based on Google Scholar with the exuberant title ‘Publish or Perish’¹¹ (PoP). *PoP* is an Open Access software that anyone can download. It retrieves and analyzes academic citations in the scholarly journals, and some other documents (e.g. books) in addition to a large number of various metrics (Hirsch's h-index, Egghe's g-index, Individual h-index, Age-weighted citation rate (AWCR) and AW-index and so... (Annex 1) somewhat similar to the metrics used in the *WOS* and in *Scopus*.

The truth is that at this stage, the tool known as *PoP* gives a much better account of the various forms of communication in the SH than the *WOS* and *Scopus*. Furthermore, its chief advantage is that it makes it possible to intervene in calculations about the index, in selecting the publications that are considered to be truly indicative of the scientific activity of an author, eliminating publications duplications or references that are not really pertinent, and in proposing a whole variety of indexes that correct for instance the h index¹².

However many problems have been pointed out when using *PoP*, such as the homonyms, the absence of a large part of European production, for instance, from Hungary, the time offset between the results displayed in *Google Scholar* and the results displayed in *PoP*, the inclusion of self-citations in the statistics and the irregularity of the results. For the last point, I (Kosmopoulos, 2009) demonstrated that, strangely, one may find insignificant books while major contributions that have been awarded are not included. At the same time, not all of the publications necessarily appear, even if they belong to the same journal and the same issue.

¹¹ <http://www.harzing.com/pop.htm>

¹² In this way, the g index (Egghe's g-index) is calculated on the same principle as index h, but gives more weight to frequently-cited articles. Another index (Individual h-index) corrects index h with the average number of authors per article, in order to make it possible to compare output between the disciplines in which the procedures for identifying the authors of an article are very different. A further feature of index h is its proper measuring according to the number of years of publishing activity of the author (Age-weighted citation rate), which entitles it to compare people who are in different phases of their scientific careers. Index h can also be calculated by assigning more weight to recent articles, giving a higher score to people who are still productive (Contemporary h-index) (Kosmopoulos, Pumain, 2007)

Conclusion for *GS* and *PoP*

Even if it is true that for most of the disciplines in the SH, *GS* and *PoP* produce much more results by including a wider production, it would need a significant and systematic overhaul for these tools to become an incontestable reference for the impact studies. The relative randomness of the displayed results and the fact that we do not have any control on the coverage, do not guarantee the necessary scientific credibility for assessment.

More generally, because of the opacity of the sources and the incomplete coverage, the information extracted from *Google Scholar* and analyzed by Anne-Will Harzing in *PoP* cannot yet guarantee a reliable evaluation.

2.6 Evaluation activities of national research systems: The Case of France

In 1987, *The Association for the Measurement of Sciences and Technology* (Courtial, 2003) was created, followed in 1990 by *The Observatoire des Sciences et des Techniques (OST)*. As it was happening in the United States with the *NSF*, the idea was to provide an independent structure able to produce metrics on the scientific research in France.

In France and in numerous countries, there is a strong tendency to apply this model in Humanities and Social Sciences. Actually there were several attempts between 2007 and 2009 of the CNRS (France) to use the metrics of the *WOS* or *Google Scholar/PoP* for evaluating the researchers and the units in SH. However in 2010, the finding was that the results were not relevant and the request of including these metrics in the annual quantitative reports was abandoned.

In 2007 a new Agency was founded: *L'Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur (AERES)* for assessing the national Institutes and the research Units. In July 2008, the *AERES* published lists of SH journals, relying on the *ERIH's* lists, others established by the CNRS, and commissions of experts. At the beginning about 6,000 different titles were inventoried in these lists. The following years the lists have been updated and sometimes the content has been radically modified. In 2012, the total number of journals listed by the *AERES* reached 8,703 titles (8,746 in 2013), covering all the disciplines in the Social Sciences and in the Humanities¹³. As well as for *ERIH*, no indicator accompanies the lists.

¹³ In the year of 2014, the *Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur* (HCERES) has been substituted to the *AERES* and joined by the *OST* : <http://www.hceres.fr/>

Therefore, since 2010 the quantitative assessment for the SH in France is simply based on whether or not the articles belong to journals that are indexed by the *AERES*.

Conclusion for the *AERES*' lists

The total amount of Journals in *AERES* lists reaches 8,703 titles, covering all the disciplines in the SH while at the same time, the *WOS* (*AHCI* and *SSCI*) indexes only 4,097 journals, that it is to say, more than double journals listed by the *WOS*.

However the *AERES* only provides lists of selected scholarly journals but no bibliometric tools. As for *ERIH* one can check whether a journal is included or not in *ERIH* lists, but no statistical indicators would complete this information.

As for *ERIH*, the *AERES* lists are public and in Open Access¹⁴.

¹⁴ <http://www.aceres-evaluation.com/>

3. The accuracy of the bibliometric data for the SH

In Section 3, I study the validity of the sources described above and scrutinize several surveys which have been done on the different sources developed in the previous Section.

We have identified above the “bibliometric” databases, tools or lists generally used in Europe to produce information or metrics about the articles published in scientific Journals in the domain of the SH.

At this stage, an in-depth review of the content of these existing tools becomes a necessity so that one can appreciate whether these bibliometric tools and lists do objectively represent the Journals in the domain of the SH and if their coverages are sufficient to produce valuable metrics or qualitative analysis if possible. In other words, the main questions that should be asked are: What does the content of the different ‘bibliometric’ resources objectively cover? What value can we give to statistical results on which these bibliometric results are based?

We propose here to evaluate the above studied bibliometric resources (*WOS ACHI-SSCI*, *Scopus*, *ERIH*, *AERES*).

3.1 The *JournalBase* project

In 2009 M. Dassa and I presented for the first time a comparative table of contents of the above databases that list the journals in the Social Sciences and the Humanities (SH) named *JournalBase*¹⁵ (Dassa, Kosmopoulos, 2009). The study covers the *AHCI* and the *SSCI* of the *Web of Science* and *Scopus*, as well as the *European Reference Index for Humanities (ERIH)* and the lists of the French *Agence pour l’Evaluation de la Recherche et de l’Enseignement Supérieur (AERES)*.

The objective of this work was to provide information about the contents of these databases at a time when bibliometric tools are raising much discussion about their application in the field of the SH. The research was carried out in 2008 with the financial support of the CNRS in

¹⁵ <http://journalbase.cnrs.fr/>

France. At that time, with some 10,000 different journals, this was an almost exhaustive overview of the wealth of publications in the Social Sciences and the Humanities, at last made available in a database, adopting near the same nomenclature for classifying the journals according to their disciplines as the one used by the *European Science Foundation* (Annex 2). The multiple assignments reveal the multidisciplinary nature of the journals, which is quite frequent in SH, but also sometimes the inconsistency of databases that have not been corrected.

JournalBase is bilingual and offers to the whole international scientific community in SH, laboratories, researchers but also European and international scientific institutions (ESF, CE, OST in Canada), producers of scientific information in the broad sense of the term, commercial publishers, publishers of open access journals, IST experts and the national assessment authorities and so on... to have free, unrestricted access and to be able to use these data¹⁶.

JournalBase carries on until today. Since 2009, we update the databases of *JournalBase* each year. It allows us to study in detail the coverage of the different resources available for quantitative assessment in the SH concerning the journals in terms of cross-comparisons, languages, countries, disciplines and so on.

Some of these studies that we had the opportunity to present in several symposia or to publish would be of interest for the present report. Especially the comparison of the coverage of the different bibliometric databases such as *WOS*, *Scopus* with the European lists *ERIH*.

3.2 A comparative study of the *WOS*, *Scopus*, *ERIH*, *AERES*

One relevant indicator coming from the cross-comparison of the four databases and lists is that less than 10% of the total number of Journals is common to the four resources (*WOS*, *Scopus*, *ERIH*, *AERES*). In other words, we found that only 1,333 Journals are common to all the sources on a total of 13,525 Journals indexed in *JournalBase* for 2012¹⁷. No particular change has been noticed in the latest update in 2013. It is then clear that none of the bibliometric resource

¹⁶ The feedbacks given by *JournalBase* users indicate that the website is used for comparative information on listings, but also for all the practical information provided on each journal. It is also used for the bibliography of assessment reports. For researchers and PHD students, *JournalBase* is a source of information and links to journals that concern their discipline. Early statistics show that in 2012 *JournalBase* spread worldwide. After France, the website is now visited in Tunisia, Chili, Belgium, the United States, Spain, Algeria, Morocco, Switzerland and Canada.

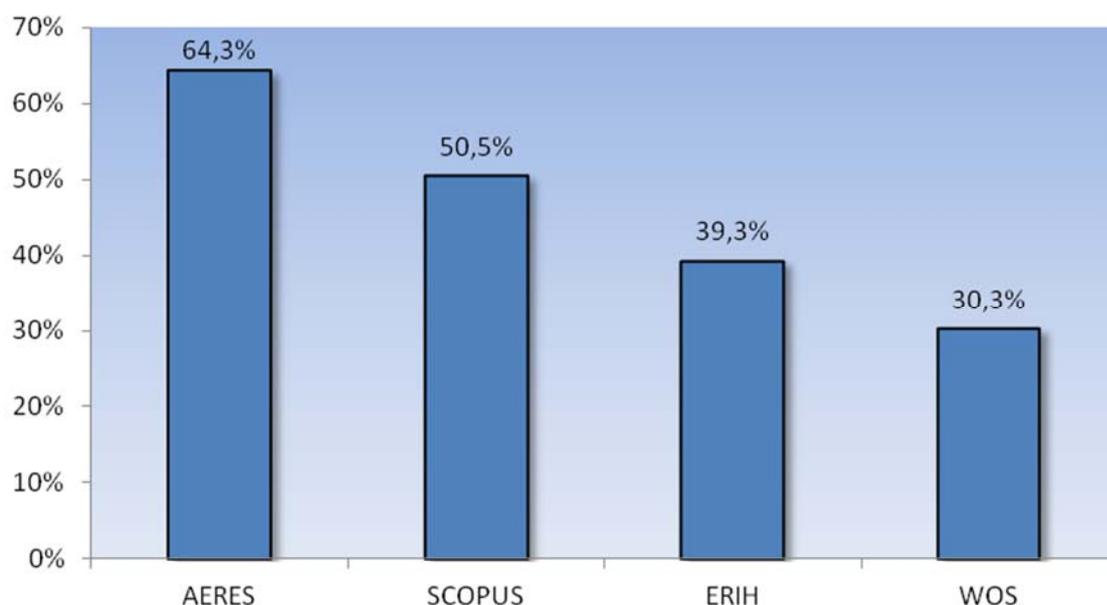
¹⁷ For instance, only 24 French journals in Humanities have been identified as common to the four sources and 15 in Social Sciences for the three sources if we exclude the *WOS-AHCI* and *ERIH*

mentioned above is in position to provide relevant information or metrics for journals or journal's articles if it is used separately. If only one resource is used for the evaluation, it would never give clear and objective information of the impact or of the quality of the publication.

During the different updates the results remained quite the same, in terms of percentage.

As shown in Figure 1, the highest percentage of coverage in *JournalBase* is the national French list of Journals produced by the *AERES*. However the result is partly biased if we consider that ERIH covers only the Humanities which is not the case for the *WOS* (*AHCI* and *SSCI*), Scopus and the *AERES*.

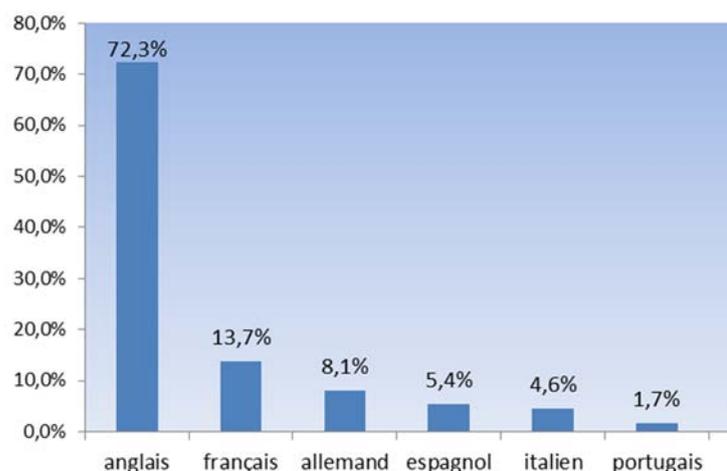
Figure 1: % for each resource/total number of journals in *JournalBase* (2012)



Source: Kosmopoulos C., Dassa M., *JournalBase – The Journals in SH and their indexing in the WOS, Scopus, AERES and ERIH*, Conference at INED, Paris, 13 October 2013

Considering the language of publication, figure 2 shows that English is the first language used in the journals indexed in *JournalBase*. French is the second language, before Spanish and German (even if English remains far ahead). Further investigations demonstrate that French journals are rarely written in another language (only 11.4%), unlike German journals, of which 45% are not in German, and 37% in English. We will come back later to the linguistic issue which is specific to a number of disciplines in the SH.

Figure 2: Distribution of the journals by language of publication (JournalBase, 2012)



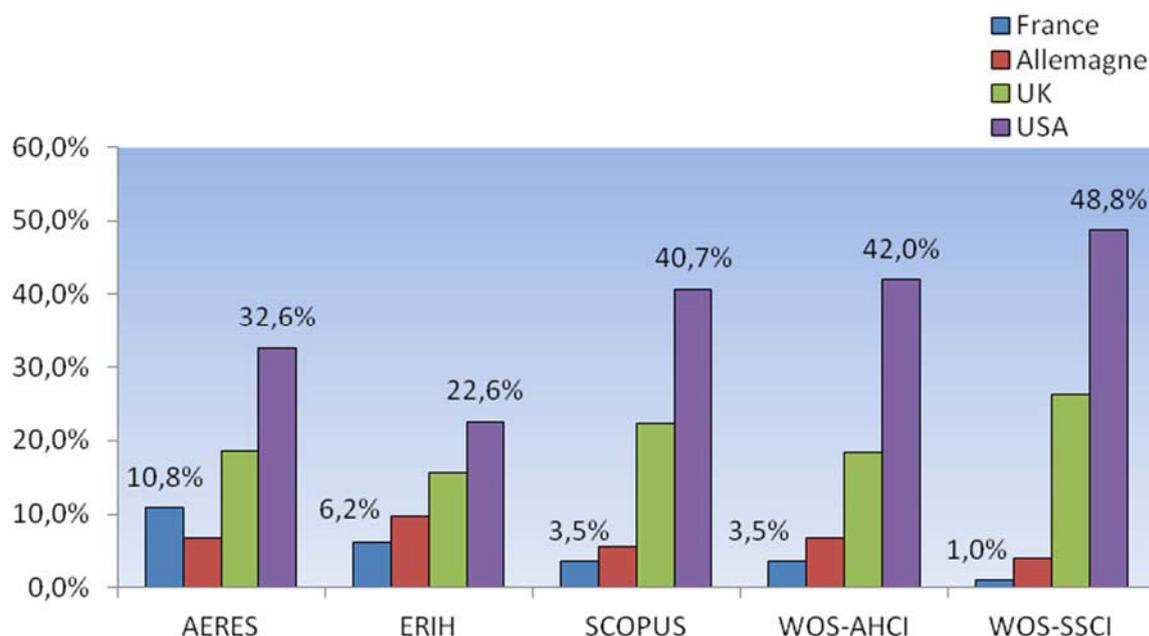
Source: Kosmopoulos C., Dassa M., *JournalBase – The Journals in SH and their indexing in the WOS, Scopus, AERES and ERIH*, Conference at INED, Paris, 13 October 2013

Figure 3 shows the distribution of the journals per source and per country of the publisher confirming the hegemony of the Anglo-Saxon production (mauve and green) in the majority of the bibliometric resources. *ERIH* seems to be more equilibrated in terms of geographical distribution.

It clearly appears that the commercial databases *WOS* and *Scopus* index a majority of Anglo-Saxon Journals, definitively more than 60% of the Journals, and even close to 75% for the *WOS-SSCI* resource.

The *ERIH* lists aims to bring out the European journals and to overcome their lack of representativeness in the commercial databases *WOS* and *Scopus*. It might be the reason that only 38.3% of the Journals published in the UK and the United States are indexed in these lists for the Humanities. The national French list produced by the *AERES* covers more than 50% of Anglo-Saxon journals. This proportion has slightly increased since 2010 with the inclusion of new disciplines.

Figure 3: Distribution of the journals per source and per country (*JournalBase*, 2012)



Source: Kosmopoulos C., Dassa M., *JournalBase – The Journals in SH and their indexing in the WOS, Scopus, AERES and ERIH*, Conference at INED, Paris, 13 October 2013

Concerning strictly the Humanities, it is interesting to notice that the coverage of *ERIH* and the *WOS-AHCI* do not match at all.

Conclusion of section 3

As it has been shown on the basis of *JournalBase*, none of the international bibliometric tools (*WOS-AHCI*, *WOS-SSCI*, *Scopus*, *Publish or Perish*, *ERIH*, and *AERES* for France and Europe) cover the entire production of the journals in SH and have the capacity to provide objective indicators on their results, more specifically on the articles' impact or on the quality of a journal. Actually less than 10% of the contents overlap between the bases. This means that none of them are completely reliable and exhaustive. To get an idea of the quality and the impact of a scientific article, each of the resources should be taken into account. *Google Scholar* and *Publish or Perish* could also be used in that case complementarily.

The *ERIH* lists can be considered as the most reliable and European base, but are limited to Humanities. *ERIH Plus* for Social Sciences is presently under development and cannot yet be used.

Moreover, according to a study conducted by M. Dassa in France dedicated to the production of the CNRS, it has been demonstrated that *JournalBase* with its near 14,000 SH' journals covers only a part (50% or less?) of the journal's publication in the SH. Indeed, M. Dassa shows that for the 2,220 articles published in 2011 by 1,762 researchers in the SH at the CNRS, only 693 journals were indexed in *JournalBase* whereas 635 journals were not covered by *JournalBase* (Dassa, 2013). These results prove that no exhaustive and valuable list of the SH's journals actually exists.

4. What are the specificities of the SH?

In Section 4, I synthesize several publications about the practices and the supports involved in the dissemination of the SH's outputs and produce results regarding to the different fields included in the domain of the SH as suggested by the ERCEA.

In the previous section, I explored the bibliometric tools or lists for all journals while the SH include several disciplines with very heterogeneous practices and types of productions. Furthermore, applying bibliometric tools as practiced in the natural sciences, based on journals alone, gives a very incomplete vision of the scientific production in SH and generates erroneous results on the number of publications and citations. Some studies have also emphasized the distinctive nature of the modes of distribution of knowledge in SH, books being able to cover 50% of the scientific outputs of certain laboratories and certain disciplines. The diversity of research methods and subjects of study, the fact that these are inseparable from the social and linguistic context in which the research is carried out, render impossible any easy and objective international comparisons of the readership of the publications. To support this idea I further explore the specificities of the SH.

4.1 The types of production according to the different fields of the SH

The research in France is developed in the universities but mostly in collaboration with different institutions specifically dedicated to the scientific research, such as the Centre National de la Recherche Scientifique (CNRS) which is the largest institution for scientific research in France including around 30,000 people.

The CNRS is composed of 10 different institutes¹⁸ covering all the domains of scientific research. One of these institutes exclusively concerns the SH; it is named the Institut des Sciences Humaines et Sociales (InSHS)¹⁹.

¹⁸ <http://www2.cnrs.fr/presse/communique/1717.htm>

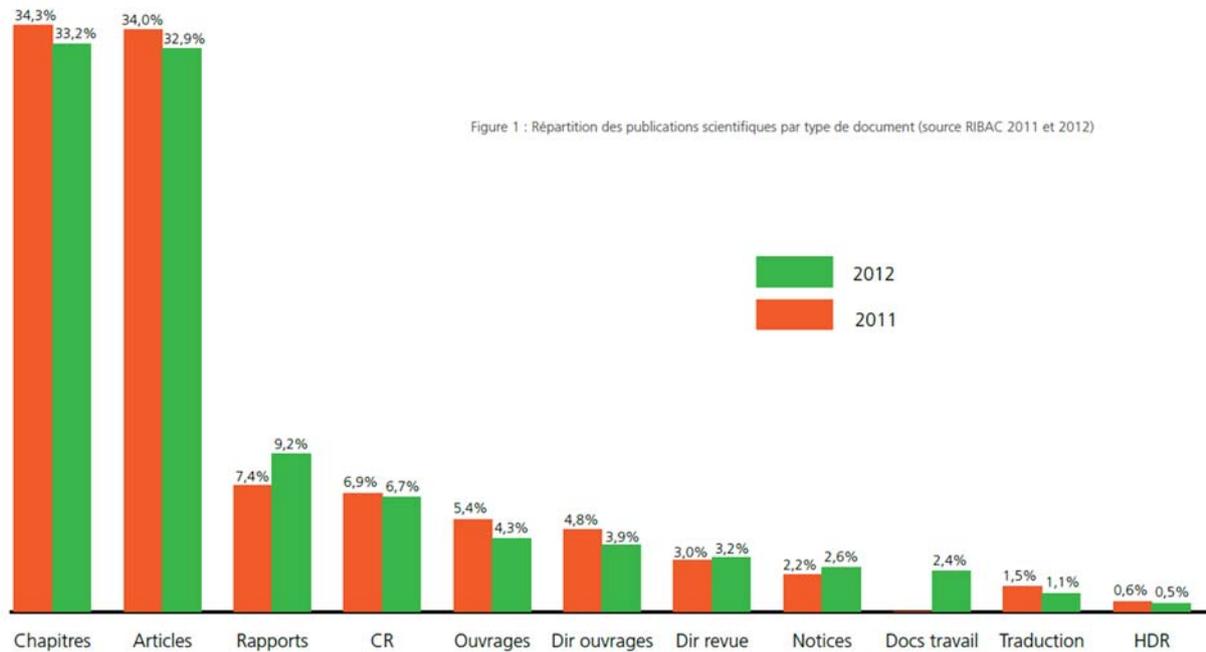
¹⁹ <http://www.cnrs.fr/inshs/>

In 2010, the InSHS developed a new tool/database named ‘RIBAC²⁰’ in order to observe in detail the production and the activities of the researchers in its domains. Since 2012, all the researchers of the InSHS (employed by CNRS) have to fulfill *RIBAC* every year. It provides relevant information about the specificity of the SH in comparison to other disciplines such as Biology and Medicine in which most of production refers to articles in peer-reviewed journals.

In 2014, M. Dassa shows the distribution of the scientific publications in the SH by type of document (Dassa, 2014) that we reproduce in figure 4.

²⁰ RIBAC: Recueil d'Informations pour un oBServatoire des Activités de reCherche en SHS: <https://www.ribac-shs.cnrs.fr/>

Figure 4: Distribution of the scientific publications by type of document (RIBAC, 2012)



Source: Dassa M., janvier 2014, *La Lettre de l'InSHS*

The study covers the practices of 97.9% of the researchers of the CNRS in the SH involving a total of 1,798 persons. Regardless of the disciplines, figure 4 shows that 2/3 of the publications in the SH concern chapters in books (33.2%) and articles in peer reviewed journals (32.9%). Books and reports cover around 20% of all the above categories of "publications" identified in *RIBAC*.

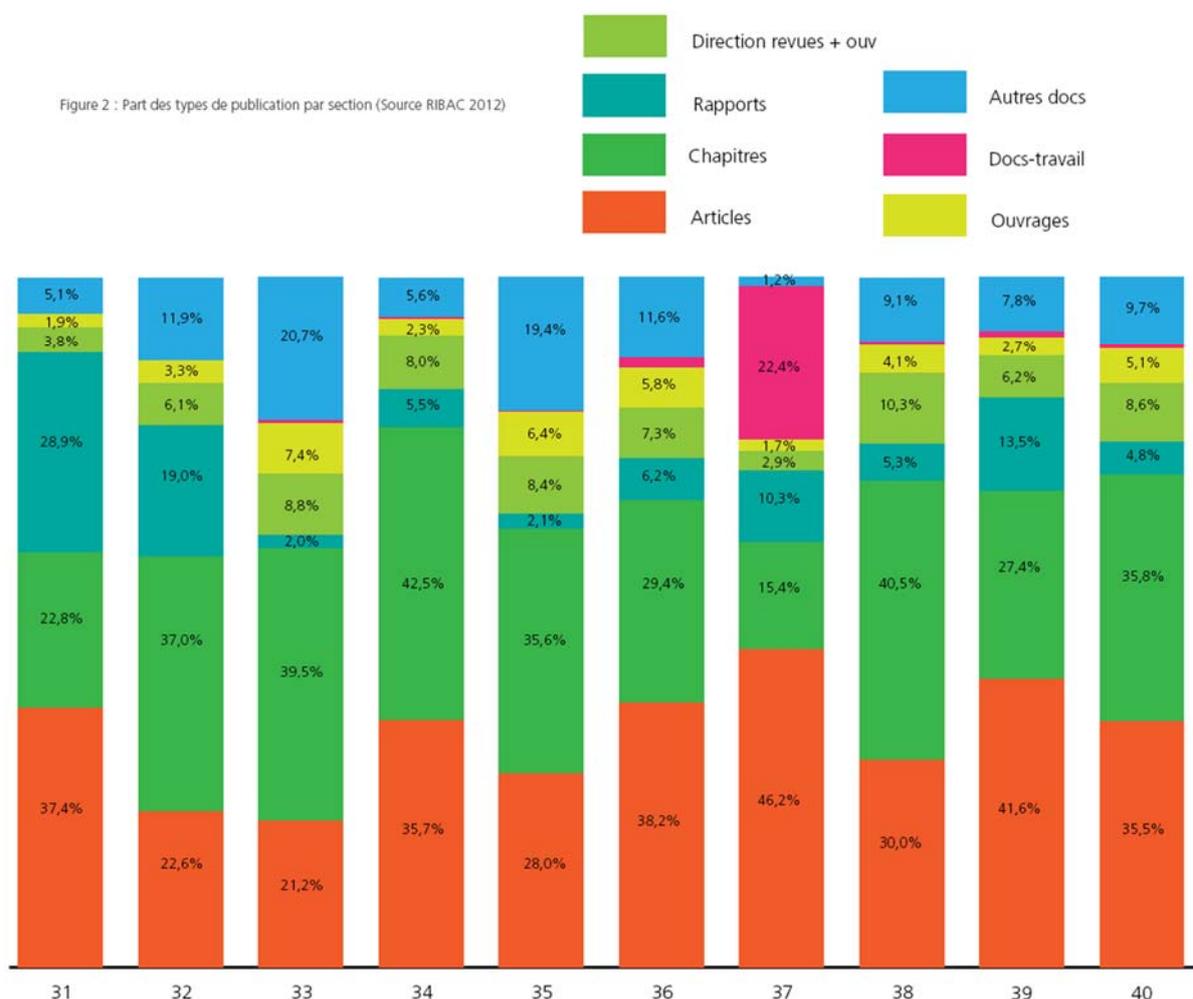
The other categories reproduced in figure 4 are: book reviews, journal's editor (special issues), working papers, translations, scholarly productions (PhD...). It should be noted that a book takes a long period to be achieved, even if it does not mean that for 10 years, for instance, the searcher will remain silent. During this 10 year's period the activity of research is naturally going on and the dissemination of the outputs might have been presented in conferences or published in articles.

To understand the variety of the distribution through the different disciplines, it has to be known that the InSHS of the CNRS is composed of 10 sections ²¹ (Section 1 is mostly attributed to the Institute for Environment).

We reproduce below in figure 5 the distribution by sections (of the InSHS at the CNRS, France) of the articles, chapters, reports, working papers, and books based on the statement of 1,798 researchers in the SH.

²¹ <http://www.cnrs.fr/inshs/presentation/sections.htm>.

Figure 5: Distribution through the different sections of the InSHS (SH) of the articles, chapters, reports, working papers, and books



Source: Dassa M., janvier 2014, *La Lettre de l'InSHS*

Caption: Section 31: History, Ecology (section shared with the Institute for Environment) - Section 32: Archeology, medieval History – Section 33: Modern History – Section 34: Linguistics – Section 35: Philosophy, Classics, Arts, History of Art – Section 36: Law - Section 37: Economics – Section 38: Anthropology, Ethnology, Sociology of Religion – Section 39: Geography – Section 40: Political Sciences

As shown in Figure 5, journals articles (red) represent more than two thirds (46.2%) of scientific production in economics (section 37) and around 20% in modern history (section 33). Most of the SH produce a high number of chapters of books (green); linguistics (section 34) has the highest percentage (42.5%); economics (section 37) looks like an exception with only 15.4% of its total production. The reports (blue-green) look significant in section 31 including ecology

and reach near one third of the total production, whereas in disciplines such as modern history (section 33) or philosophy and classics (section 35), they represent only 2%. In general, books or ‘books edited by..’; occupy a quite relevant place in the SH production, except again for economics (section 37) and ecology (section 31). In modern history (section 33), books represent over 15% of the production and for the rest of the disciplines the average is around 10% to 12%. With their “working paper”^{s22} (fuschia), economics (section 37) show once again their particularity.

The blue color corresponds to diverse documents, including book reviews that are considered to be essential in History, but it also concerns translations as well as scholarly works such as theses. Economics demonstrates its specificity in the production of diverse documents with a very low % (1.2%).

Connecting *RIBAC* and *JournalBase* (Dassa, 2013) makes it possible to give some indicators about the articles published in peer reviewed journals. In general, it has to be underlined that regardless of the disciplines, only half of the articles published by the researchers in the SH at the CNRS are indexed in *JournalBase*.

The study of the 2,220 articles stated in *RIBAC* for 2012 reveals that the Economists (with 75% of their articles) and the Linguists (with 66% of their articles) are the most represented in *JournalBase*. This can be explained by the fact that the Economists are quite well indexed in the *WOS-SSCI* (Social Sciences) with more than 45% of their articles. Philosophy and Classics are also in a good position with 30% of their articles stated in *RIBAC* indexed in the *WOS-AHCI* (Arts and Humanities).

In contrast, the scientific productions in History and Anthropology/Ethnology are less indexed in *JournalBase*. Only 40% of the articles belong to the journals listed in *JournalBase*, indicating that 60% of the journals in which the scientists of those fields publish, are not indexed in *JournalBase*. As a matter of fact very few journals from such disciplines are present in the WOS (Annex 3).

²² In economics working papers are used as pre-prints.

4.2 The pre-eminence of presentations at congresses

Apart from the type of documents mentioned in figure 5, one needs to acknowledge the relevant amount of communications to symposia or congresses. According to the statement of 1,798 researchers, 90% of them have been communicating in at least one symposium during the year 2012. About half of the communications were invited conferences (Dassa, 2014)²³.

The 7,713 communications to symposia (compared to the 2,220 peer reviewed articles) recorded in *RIBAC* for 2012, indicate that this type of production is the most important vector for dissemination in the SH and that it should also be taken into account in a quantitative analysis.

4.3 The linguistic diversity

The issue about the language always resurfaces in the SH. In most of the so-called “Hard” sciences, the current language of the publications is English and it makes them easier to compare and replicate on already standardized objects of research. This of course cannot be the case for cultural studies or localized analysis in SH where research objects are frequently embedded in different societies and languages.

The study of *JournalBase* - listing nearly 14,000 different journals in SH - identifies 49 different languages used in the articles, besides some journals qualified as “multilingual” (Annex 4).

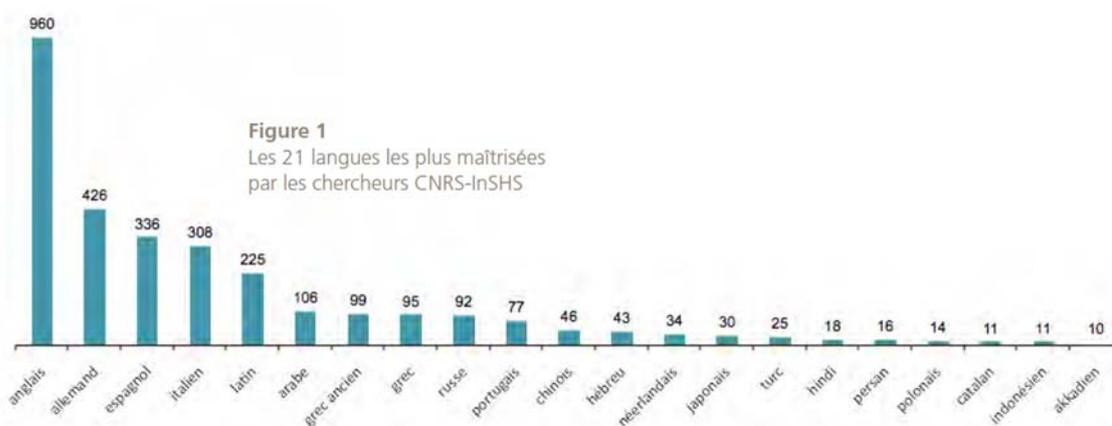
In 2010, another survey involving 976 French researchers of the CNRS of all the SH’s disciplines, indicates that they know in total 197 different languages (Dassa, 2012). It demonstrates the wealth of knowledge on the subject and how crucial the question of language remains in these fields. Some disciplines are basically related to the lingual aspect. For instance, certain sections in History, Anthropology, Literature, Epistemology or Philosophy of Science etc. cannot be translated into English precisely due to the language as specific subject involved in these studies. Moreover in those cases, an English translation would deeply alter the meaning of the current analysis.

²³ These last results should be taken with great caution. Indeed the status of ‘invited conference’ is not always well understood by the researchers in the SH. Practically it means that all expenses have been afforded by the organizing committee.

On the contrary, in Economics, Linguistics, and Psychology, the majority of the articles/publications are in English and generally quite well-represented in the Anglo-saxon databases such as the *WOS* and *Scopus*.

A classification of the first 21 languages identified in *RIBAC* as the most well-known (Figure 6) matches with the fourth first languages listed in *JournalBase*.

Figure 6: The 21st languages more practiced from a sample of 976 SH's researchers at the CNRS (France)



Source: Dassa M., mai 2012, *La lettre de l'InSHS*

As we can observe in Figure 6²⁴, the most well-known languages - as stated in *RIBAC* - are: English, German, Spanish, Italian, Latin, Arab, Ancient Greek, Modern Greek, and Russian.

The languages mostly used as listed in *JournalBase* are in order: English, French, German, Spanish, Italian, Portuguese, Dutch-Flemish. A detailed description of the 49 languages detected in *JournalBase* is available in the Annex 4.

²⁴ French is of course not included in the figure 6.

In addition to these results, *RIBAC* analysis shows that English is the first language used in articles (46.1%), then French (39.6%) which is not surprising for a French institute..., 5.1% in German, 3.2% in Italian, and 2.7% in Spanish (Dassa, 2013).

If this exercise would be done in other European countries, would the proportion of the publications in English, native, and other languages remain the same?

Conclusion of section 4

The results of Section 4 strongly support the hypothesis that the disciplines in the SH radically diverge from the natural sciences in terms of productions and practices. In contrast to Medicine and Biology for instance, the SH generate a wide diversity of publication supports, in which the articles published in International peer-reviewed journals correspond in average to only 33% of the total of their production. Moreover the SH globalize many disciplines with recognized various practices. The Economists do mostly publish in scientific journals, in English, but this not the case with the Historians who in priority produce books and books' chapters. Our findings reveal that regardless of the differences between the disciplines, 2/3 of the 'essential production in the SH is related to the articles and book's chapters including many reports or working papers for some sub-disciplines; 1/3 is related to other types of production. As it has been shown in this section, the SH, regardless of the disciplines, usually do produce a broader content such as: reports, translations, book's reviews, working papers, scholarly work, communications in symposia, but not only...

As this exploratory study demonstrated, the SH are not homogenous. Each discipline carries a specific background which makes it difficult to examine them globally.

The issue about the language is one of the specificities of certain disciplines as well as an undeniable and irreplaceable richness for the Humanities. In that case, classical metrics such as the calculation of the impact by the number of citations is simply not suitable. The definition of 'quantity' and 'quality' should be in that context reformulated in regard to the respective field.

A general consequence of the above remarks is that the usual bibliometric tools applied in the other fields of science are not adapted either to the variety of production of the SH or to the nature of the studied objects. Combined with the fact demonstrated in the previous section that neither the *WOS* or *Scopus* or *ERIH* or a national list as the *AERES* is in position to produce an objective overview of the production in the SH, one would rather be tempted to say that the SH, because of their specificities may have a leading role to play in sustaining and developing new

models of quantitative assessment. The development of the new technologies and new practices enable us to build a variety of better adapted tools for different scientific contexts.

A different approach of these questions, including historical but also new practices, could drive the ERC to implement new tools less general but much more valuable for decision-making. The next section will enlighten us about new practices emerging with the modern Information and Communication Technologies (ICT).

5. New practices: Science in Open Access

The modern Information and Communication Technologies (ICT), and more specifically the web, have brought about profound changes that cannot be ignored in communication on a global scale, and therefore in all scientific practices now depending on it. Therefore, we should pay close attention to these evolutions that may radically modify the scientific discovery process.

The aim of Section 5 is to understand how far the emergence of the ICT could influence the nature of the scientific construction of the truth in the SH and transform their manners in producing results. The Open Access Initiative has been developed within the scientific community since the 90's. The OA mechanism has deeply influenced the European policy as it is proved with *Horizon 2020*.

5.1 The emergence of Open Access (OA) practices

In terms of content and practices, "Open-access literature is digital, online, free of charge, and free of most copyright and licensing restrictions....When copyright holders consent to OA, they usually consent in advance to the unrestricted reading, downloading, copying, sharing, storing, printing, searching, and linking of the full-text of the work. Most authors choose to retain the right to block plagiarism, misrepresentation, and sometimes commercial re-use. They authorize all the uses that are needed for legitimate scholarship, including those required by the technologies that facilitate online scholarly research" (Suber, 2004).

The international Open Access movement was made official by the *Budapest Open Access Initiative* in December 2001.

5.2 The different models of OA's scientific content

A first category would be **Open Access Publishing**, which includes all the Open Access peer-reviewed Journals answering to the OA criteria mentioned previously. At present, 10,633 academic OA journals are listed in the *Directory of Open Access Journals (DOAJ)* –hosted by the National Library of Sweden²⁵. Recently different institutions started to make books

²⁵ <https://doaj.org/>

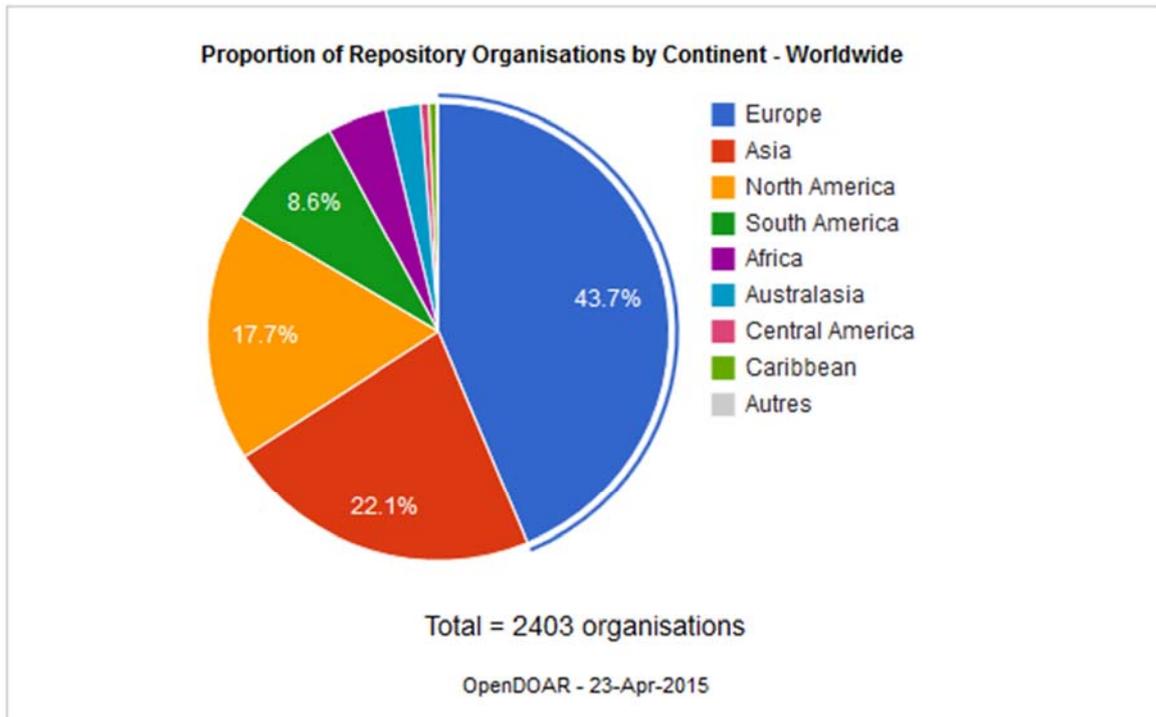
available in Open Access. *OAPEN Library*, for instance, is an Open Access publishing initiative for the Humanities and Social Sciences, started in 2008 and consisting of 31 European university Presses²⁶. The goal is to create the largest freely available collection of current peer-reviewed books in European languages in various fields of SH.

The second category -which has existed for 25 years- concerns the Repositories, also called **Open Archives**, that are being deployed in research and learning making use of full-text articles, across a range of different types (subject, national, regional, institutional, project, lab and so on.) in which the authors self-archive their pre- or post-prints from a non-OA journal or other documents as we will see further on. *Horizon 2020* promotes the self-deposit in Open Archives. The scientist is advised to deposit his or her final peer-reviewed manuscript (not the pre-print) after a certain embargo period (if it is imposed), established in order to allow publishers to recoup their investment. Presently over 2,600 academic open repositories are listed in the Directory of Open Access Repositories (*OpenDOAR*). Over 1,450 (43.7%) research repositories are run by European institutions²⁷ (Figure 7).

Figure 7: Proportion of Repository Organisations by Continent -Worldwide

²⁶ <http://www.oapen.org/home>

²⁷ Including different use case: Institutional repositories, Discipline-oriented repositories, Interdisciplinary, cross-disciplinary and cross-sectoral repositories, *National Repositories* such as HAL in France. HAL is a large national repository for research outputs with enriched metadata.



Source: *OpenDOAR*

In those repositories non-conventional productions such as working papers, theses, or more recently data, code ...are now being largely accessible online. As an example a well-known code repository *GitHub* hosts over 22 million repositories of code in an OA environment.

The third category of scientific content in OA would include new collaborative models of Open Access based on the web 2.0, like the **wikis** and **blogs** now spreading quickly through the scientific community. As an example, the so-called “Carnets de recherche” from *OpenEdition* in France exceed 1,000 blogs²⁸.

Concerning the last two categories, it has to be mentioned that they are partly peer-reviewed (e.g. pre-prints, academic works as theses and so on in Open Archives).

²⁸ <http://www.openedition.org/catalogue-notebooks?page=catalogue&pubtype=carnet&lang=en>

5.3 The services developed with OA

In addition, OA services have been introduced within the web environment that include:

- **Social media:** social networking through collaborative websites, blogs, tags, social bookmarking etc
- **Alerting:** for example, the RSS feeds allow readers, of a website, to track updates on the site as soon as they become available
- **Indexing tools**²⁹: which enable linking publications on similar issues
- **Data Mining:** extraction of significant data for re-use

5.4 The “Open Sources”

Apart from the contents or the data, one should not overlook another category: the so-called “Open source” software developed by the scholarly community which constitutes an important aspect of the Scientific Open Access issue, especially in the SH where such software is essential to make it possible to implement Open Access documents, and repositories. Many of the open source tools in science are combined with the Open Access protocol that is presented in section 5.5.

One single example is *Zotero*, a bibliographical open source manager developed by the scholarly community that automatically senses content in the web browser. With one click, one can grab all the metadata of any document online (article, book, web page, report, media, communication at a symposium, etc.) distinguishing the type of document and collecting them in a database that can also be shared with others. *Zotero* is also a network of scientists such as *Researchgate* or *Academia* (Annex 6) which enables it to exchange with known or unknown people on common subjects of research. But the difference is that *Researchgate* and *Academia* belong to private companies...and do not respect the interoperable protocol.

5.5 The interoperability of the data

To proceed efficiently, the OA movement introduced the notion of ‘interoperability’ between the documents. The objective of *Open Archive Initiative Protocol* is to render the metadata on documents online interoperable, in order to permit the transfer of data between the different

²⁹ E. Garfield was a precursor in the development of such tools before the Internet era. In respect of a defined protocol the indexing tools can today be made automatic. There are many types of *indexes*, from cumulative indexes for journals to computer database indexes.

servers respecting the defined protocol. The principle for OAI-DC or OAI-PMH is simple: the data are deposited on the server in a storage centre and made accessible to all OAI “harvesters” who navigate on the web. It is made accessible through any website personal or institutional such as European Repositories.

For instance the protocol OAI-DC³⁰ is being used by the European project *OpenAIRE*³¹. *OpenAIRE* is an online network collecting the EC funded publications, but not only³².

OA bibliographical tools like *Zotero* do work on the basis of OAI-DC Protocol as well. Building a bibliographical collection with *Zotero* or a Repository of Repositories presupposes basically the same infrastructure and the same system to automatically track the metadata. By combining a huge amount of information, OA protocol allows developing collections of Big Data and makes it possible to provide cross-comparative studies, statistical and semantic analysis, as well as cloud computing (Annex 7).

The increase of OA data modifies radically the traditional environment and practices of scientific research. A practical effect of these new developments is the increase of interdisciplinary, particularly in some disciplines of the SH (Geography, Epistemology, Linguistic, Sociology...). It facilitates moving the frontiers between the disciplines, not only within the SH but within the whole scientific system.

Such approaches allow tremendous changes and promises for researching, capturing and re-using the information at a large scale. With the ICT, new forms of scientific communication and dissemination surface and force us to rethink the practices in scientific research as well as the historical scientific publishing system and the related methods of professional evaluation. Concerning the assessment exercise, this new context of producing scientific knowledge constitutes a new way in releasing metrics.

³⁰ *Zotero* is regularly improved by scientific volunteers from all over the world. The plugin uses the OAI-DC protocol. OAI means “Open Access Initiative” and DC “Dublin Core”. DC is an open access schema that describes any kind of web resources: <http://dublincore.org/>

³¹: <https://www.openaire.eu/>

³² Since Horizon 2020, the repository also gathers other Open Access publications, regardless of their funding. It is of course possible if originally the websites that are ‘harvested’ respect the standards of the protocol. The French national Repository HAL (CNRS-CCDS) is being harvested by *OpenAIRE2020*: <https://www.openaire.eu/news-events/openaire2020-press-release>

5.6 Altmetrics

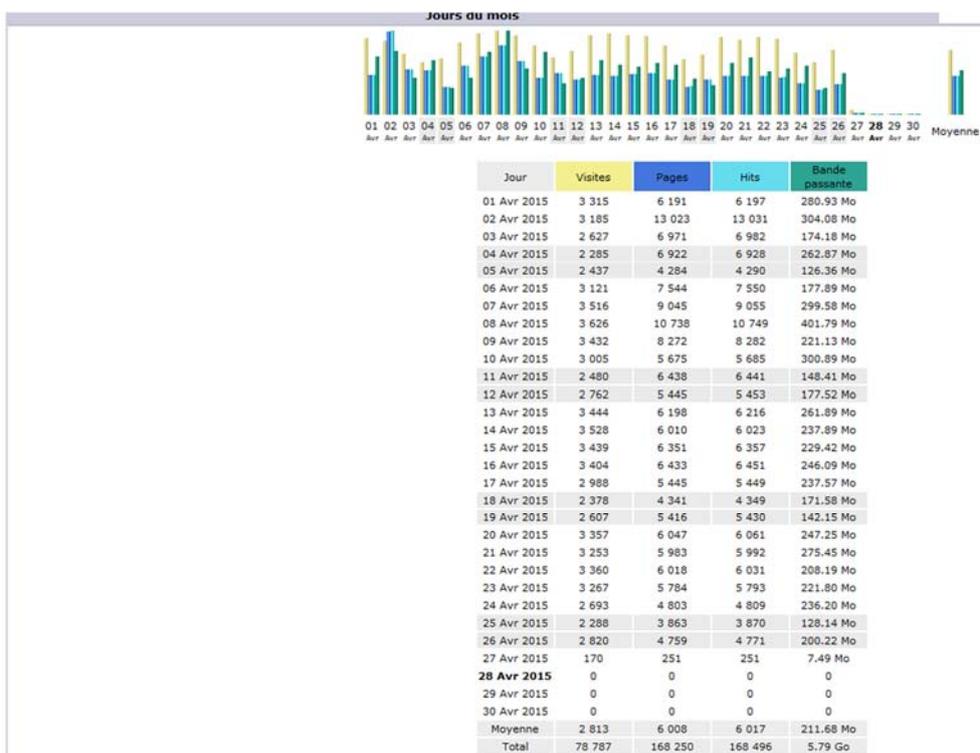
“Alternative metrics” or altmetrics: we approach here a continent recently emerged. The exploration already engages many actors, especially those involved in the question of assessment: researchers themselves, research funders, universities, research institutes....but also private publishers.

The partners of the OA movement have been the first to introduce the idea of altmetrics. In practice it has demonstrated that "OA maximizes research access, uptake, usage, impact, productivity, progress and benefits to humankind" (Harnad, Carr, Gingras, 2008). With such enriched sources, "the metrics are becoming far richer, more diverse, more transparent and more answerable than just the ISI Journal Impact Factor" (JIF). Since 2008, the altmetrics became a central question of the OA movement and several indicators such as author/article citations, author/article downloads, book citations, semiometrics, hub/authority metrics and more, have been tested in the respect of the OA protocol.

Presently, a large number of OA metrics are under study, but already new measures that become possible with online publication are the number of visits, number of views, downloads or 'hits', particular visitors, opening a new line of investigation (figure 8).

Figure 8: Example of altmetrics for an open access journal (Cybergeog)

- Résumé
- Quand:**
- Historique mensuel
- Jours du mois
- Jours de la semaine
- Heures
- Qui:**
- Pays
 - Liste complète
- Visiteurs Robots/Spiders
 - Liste complète
 - Dernière visite
- Navigation:**
- Durée des visites
- Types de fichiers
- Downloads
 - Liste complète
- Pages vues
 - Liste complète
 - Entrée
 - Sortie
- Systèmes exploitation
 - Versions
 - Inconnu
- Navigateurs
 - Versions
 - Inconnu
- Origine/Referer:**
- Origine de la connexion
 - Moteurs de recherche
 - Sites référenciers
- Recherche
 - Phrases clés
 - Mots clés
- Autres:**
- Divers
- Codes Status HTTP
 - Pages non trouvées



Caption: day, number of visits, views, downloads, bandwidth

Brody *et al.* have been prominent in showing that there is a correlation between higher downloads and higher impact, particularly for high impact papers, holding out the promise not just for higher impact resulting from open access but for the ability to predict high impact papers much earlier, not waiting years for those citations to materialise (Harnad, Brody, 2004). Actually the University of Southampton (UK) registered in 2013 more than 30,000 downloads/month (for a total of 6,000 documents archived) and the University of Liège the same amount for a total of 35,000 documents archived in their open access repository.³³

It is usual that the OA repositories do instantly display such metrics at a global level but also about a particular document. As shown in figures 9 and 10, a document is being provided with all its metadata, the full content is freely available in pdf format and the metrics calculating how many times the metadata have been consulted and how many times the full document has been downloaded, are indicated. At the same time the interoperability allows the linking of the document to all the other contributions in OA on the same subject and to the Open Archives

³³ Principes directeurs pour le développement et la promotion du Libre Accès, rapport de l'UNESCO dirigé par Alma Swan, 2013, p.26

Collections of the institutions which have collaborated for the achievement of the Phd thesis (figure 9).

Figure 9: Example of altmetrics in the French National Open Archives Repository HAL (CCSD)

The screenshot shows the HAL repository interface for a thesis. The main content area includes the title, author name (RENAUD LE GOUX), and an abstract in English. Below the abstract are keywords (AMENAGEMENT, COMMUNAUTARISME, ETATS-UNIS, GATED COMMUNITIES, IMMOBILIER, INTRA-URBAIN, SEGREGATION) and document details (type: Thesis, domain: Humanities and Social Sciences / Geography). A 'Display the full list of metadata' button is visible. The right sidebar contains several sections: 'FILE' with a PDF download icon and HAL ID; 'IDENTIFIERS' with the HAL ID; 'COLLECTIONS' listing various academic collections; 'CITATION' with a citation text; 'EXPORT' with options for BibTeX, XML, and endNote; 'SHARE' with social media icons; and 'METRICS' showing 959 consultations and 741 document downloads.

Figure 10: Example of altmetrics in the repository of the University of Southampton, UK

 RSS 1.0

 RSS 2.0

 Atom

Item Type: Article

Related URLs: <http://revista.ibict.br/liinc/...php/liinc>

Keywords: open access, institutional repositories, citation impact, scientometrics, research policy, deposit mandates

Divisions: [Faculty of Physical Sciences and Engineering > Electronics and Computer Science > Web & Internet Science](#)

ePrint ID: 266617

Date Deposited: 02 Sep 2008 17:47

Last Modified: 27 Mar 2014 20:12

Further Information: [Google Scholar](#)

URI: <http://eprints.soton.ac.uk/id/eprint/266617>

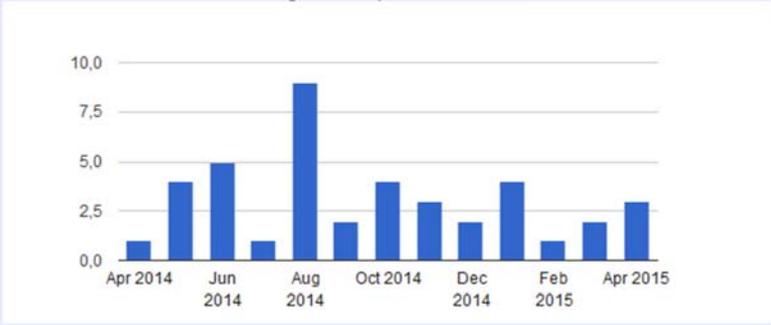
ASCII Citation  URI & RDF

Actions (login required)

 View Item

Download Statistics

Downloads from ePrints over the past year. Other digital versions may also be available to download e.g. from the publisher's website.



[View more statistics](#)

As a matter of fact, altmetrics do not have a limited definition and are still under study and development. The idea is to build tools that make it possible to obtain indicators about all various types of outputs available on the web, including books, but even more un-orthodox production such as communications at congresses, videos of presentation, web pages, source codes, reports, ppwt, blogs, as well as social media³⁴ and more... Respecting a defined protocol

³⁴ Based on the web 2.0, the social media have been described in section 5.3. They make possible to create in common and to share with others research outputs.

these tools make it possible to produce data mining and create ‘Big Data’ that can be re-used and exploited by algorithms in order to extract the necessary information or specific metrics.

The data mining has been made possible by the development of the modern technologies and the progress in the OA environment. More precisely data mining permits the extracting and analysis of significant information. This information can be re-used in order to produce indicators. The advantage of tracking all the datasets is to trace impact as well as influences that are not mentioned or biased in the classical system of citations (section 6). Data mining can reveal underlying processes and domains of research that are not being covered.

It is then not surprising that in April 2014, the website of *Zotero* announced that the Alfred P. Sloan Foundation had awarded \$360,000 to fund two years of research into altmetrics to be conducted at the University of Montreal, Indiana University, and George Mason University. The *Zotero* project’s involves the aggregation and delivery of anonymized datasets to allow their research partners in Montreal and Bloomington to compare readership across a range of metrics, including commercial databases, social media, and reference management software. *Zotero*’s feature is to put into production a preliminary public API³⁵ that returns anonymous readership counts when fed universal identifiers (e.g. ISBN, DOI), enabling bibliometric research and integration into third-party web programs....

The resulting research expected from this funded research is to improve the understanding of social media’s value in scholarly communication and shed light on the actual meaning of various altmetrics scoring systems.

Zotero’s project is an example, and many stakeholders of the OA movement are moving towards the production of alternative metrics (e.g. *OpenEdition*).

Predictably, private companies have already invested in the still confused notion of “altmetrics”. developing websites such as the reference manager system *Mendeley* (created in 2008 and sold to Elsevier in 2013), *altmetrics.com*³⁶ (created in 2011) supported by Springer,

³⁵ Application programming interface (API) is a software component that can facilitate the implementation of new functionalities into existing applications

³⁶ <http://www.altmetric.com/>

Elsevier with *Scopus*. *Altmetrics.com* tracks the social media activity (Tweets, Facebook posts, Google+, Blogs and media..).

The dangers to apply these types of automatic collectors of information are that they: 1) do not refer to the nature of the transmitted information 2) can be biased by “fashion effects” of the news (an earthquake, the destruction of an antic site etc.) and 3) function independently of any scientific control. Hence, the whole process remains very opaque in the hands of the stakeholders. Furthermore it contributes to increasing the confusion between social scientific activity/impact and social media activity. Tweets, Facebook posts, Google+, Blogs and media cannot, as yet, be considered scientific data for observing the validity of scientific contributions or knowledge. But facing the lack of quantitative metrics, the influence of *Mendeley* as well as *altmetrics.com* seems to grow up. However one should never forget that they both are based on totally uncontrolled and arbitrary data and therefore produce very partial indicators which can in no way be weighed.

Conclusion of section 5

As it has been shown, the Internet but more specifically the OA has moved forward the practices of collecting, sharing and analyzing significant information. The key to the altmetrics process is to agree on a common protocol within the scientific community that enables the connection of all the datasets available on the web. OAI-DC is one the most efficient today, but the project is far from being achieved. It still needs to design and implement new infrastructures and systems at local and regional levels. On the one hand, the challenge of altmetrics depends on a common scientific policy and strategy, and on the other hand, it implies specific technical developments at the local level to make possible the interoperability between all types of OA resources, in order to clearly identify the sources and to base the quantitative assessment on accurate and relevant scientific data.

6. Relation between quantitative and qualitative evaluation in fields of Social Sciences and Humanities

This section concludes with the information collected in this report. Next to them, a recall of the long practice of peer review is included and compared to the opportunities of quantitative assessment in SH. Depending on the disciplines included in the SH the relations between the two models may be different.

The purpose of using scientometrics in the process of assessment is to produce objective indicators of the performance of the research outputs and/or to find out new trends of research that might bring an added value to the scientific process.

Indeed, the question of objectivity in the assessment's exercise should necessarily be asked, for either the peer review or the bibliometric indicators. What is the peer review? Why is it suspicious? Why do we think that a quantitative evaluation should be included in the assessment process? Is there a risk of using metrics? What kind of metrics may be relevant for the SH?

The peer review has a long practice which runs for centuries. Although historically the beginning of this exercise depends on the discipline, it is generally accepted today that the peer review is the fundamental basis of the scientific progress. In short, the principle is as follows: the peers - experts with similar competences - are invited to evaluate through a double-blind process a submitted manuscript and to determine whether or not it is justified to publish it. It mostly concerns the articles in journals, and partly books and proceedings.

To understand why bibliometric indicators have been introduced to supplement the peer review, we firstly should return to the Prehistorical period of the quantitative assessment exercise and consider once again its origin in the citation indexes. When E. Garfield reached the idea to use the citation indexes in order to study the impact of a journal or of an article (which means to assess the validity of the scientific posture of the authors) he believed that the citation data could objectively ameliorate the observed bias of the "Old boy network" or "Fossilized groups" that may have qualified the committees of peers if these committees are not regularly refreshed with new members (Garfield, 1998).

Actually, he noticed that the peer statement may be in some cases too subjective and reproduce conservative approaches instead of identifying innovative research fronts. Indeed the bias due to the disciplinary differences can make it difficult for interdisciplinary projects and frontiers

research to be assessed by the classical means of peer review. In contrast, the development of the ICT as well as the free access to the data conduct to the increase of interdisciplinarity, especially in the SH (section 5.5)

Other biases have several times also been underlined especially in small domains of research where most of the members of a community do know each other. Indeed, even if a submission is anonymous, it is easy to identify the author and consequently influence the decision. Besides that, the novelty in method, subject or paradigm is not always accepted and can even be suspicious to peers that are not involved in similar issues or approaches. It is particularly sensitive in the SH with the development of the ICT (about new methods of production and dissemination, see section 5) and the transdisciplinary research (e.g. “Digital Humanities”).

Another point that is being added to the criticism of the peer review is the delays that are required when evaluating an article and for publishing. It is known that these delays sometimes take years. Actually the question of delay is one of the most relevant reasons that the physicists started in the early 90’s to implement the first open archives repository (*OpenArxiv*). By self-archiving a paper in a public repository the feature not only guarantees the property of the discovery, it also allows a large dissemination of the results on the web, and data mining, even before being officially published.

Next to the peer review, the purpose of the 'classical' bibliometric indicators generated by the citations indexes would be to establish the value of a project in the sense that the number of times an article or a book is cited is an indication of its usefulness and its scientific value. This question remains obviously open to discussion, as a large number of citations can also reflect controversy, refutations, or simply the effect of a fad. Another bias signaled by Y. Gingras (Gingras, 2014) is that a high number of citations do not always show an international impact. To measure the international impact, the usual metrics should be refined. For instance, the localization of the author who cites a reference should be analyzed. If an American article is cited hundreds of times within the American scientific network, it cannot be considered as having a high impact on the international community. Furthermore, by identifying the origin (local, regional?) of each co-author collaborating to an article, we could better appreciate the real extent of the impact.

If we look deeper into the question of objectivity, one should not forget that metrics are the results of algorithms built by the human brain as well as the data building, so that the question is partially solved with bibliometric indicators.

Considering the SH, the quantitative assessment cannot so far supplement the peer review, because articles cover only 1/3 of the production in the SH – even if most of the production

varies from one discipline to another – and the current bibliometric tools have been designed for articles.

More precisely, Section 3 demonstrated that none of the official bibliometric sources for the SH are exhaustive enough to deliver credible indicators. The comparison between the different sources proved that only 10% of the journal's titles are common to all of them. This study is based on the inventory of *JournalBase* which corresponds to the addition of all SH journals' titles existing in the *WOS*, *Scopus*, *ERIH* and *AERES* (France), that means in total near 14,000 different SH's journals. Moreover the cross-study with *RIBAC* brought up the fact that half of the journals in which the scientists generally publish are not even included in *JournalBase*. Few disciplines, like economics, are better represented in these bibliometric sources with an indexing of around 30 to 35% of the journals in their domain. But still, even if we consider each discipline of the SH, we cannot objectively manage to get good indicators about the impact. The data used for the bibliometric analysis are simply not sufficient and not representative of the SH's production. Considering all the other production, the data are simply not there. This can be considered as the major disadvantage of bibliometrics.

However as it emerged from this report, in the new context of OA, altmetrics are undoubtedly the most promising project as it pertains to quantitative assessment for the SH in regard to what bibliometric tools are today providing. They surely would be more adaptable and open to the wide and various production of the SH and, favor the visibility of interdisciplinary impact rather than the recurrent indicators retrieved from the different citation indexes or checklists derived from the citation indexes of E. Garfield.

On the other hand, as any quantitative data, altmetrics may narrow the definition of science and could also affect the research creativity³⁷. It is one more occasion to repeat that we should be very careful with these new developments and that even refined metrics such as altmetrics will always have to be combined with the peer review.

³⁷ Certainly all these quantified measurements should not be considered as a substitute to the qualitative evaluation, partly subjective but more subtle because of its ability of analyzing the content.

7. General conclusion

We presently have to face intricate questions concerning the quantitative assessment of the SH. How should we proceed? Should we find a solution to reintroduce the SH into the historical scientific model of assessment already applied in the natural sciences in the continuity of what E. Garfield developed, in order to match with a certain definition of scientific research? Or should we look at the question of assessment in the SH with a different viewpoint, including in our analyses the inherent diversity of the sciences and the specificities of the scientific discovery within the SH, as well as the new opportunities brought by advanced ICT tools?

This study shows that the classical approach of quantitative assessment has been biased since the 60's. The classical approach is based on articles only and the number of times an article or a journal has been cited and concluded to the "impact" of the specific document. The tools that were already available for applying this model are the well-known *citation indexes* created by E. Garfield. What has been developed further on since the 80's through the *NSF* system of evaluation, are derived products of the citation indexes, not only in their 'material' conception but also in the vision of how the scientific progress in fundamental research should be quantified. Indeed, the entire system is based on the citation indexes which were originally designed for the medical and biological domains of research, including, as well, an orientated-scientific process in terms of production and communication. Because E. Garfield referred its indexes to scientific journals, the easiest way of conducting a quantitative study was to use the available databases (section 1). Other disciplines such as Physics (Laloë, Mosseri, 2009) or Mathematics refuse to apply such quantitative metrics for assessment.

Could this system operate within the SH? The results of this report demonstrated the following assertions:

The coverage of the two well-known commercial databases (*WOS* and *Scopus*) and the lists of academics such as *ERIH* or *AERES* (France) only concern a certain type of production, that is to say, journal's articles. The comparative study generated with *JournalBase* revealed that less than 10% of the contents overlap between the bases, which means that none of them is completely reliable and exhaustive for producing indicators about the impact of journal's articles in the SH (sections 2 and 3).

To some extent *Google Scholar* or *Publish or Perish* databases provide better coverage for publications in the SH than the *WOS*, *Scopus*, *ERIH* and national lists (as *AERES* in France),

because they include books. It would certainly be useful to undertake for each discipline and sub-discipline a test of its different measurements. However, without information about the sources of *GS*, it remains opaque and cannot scientifically be validated for evaluating the research (section 2).

Even if there is a very high pressure towards the scientists in the SH to publish articles rather than any other types of publication, the SH' production cannot still be reduced to peer-reviewed articles. Regardless of the disciplines, book's sections and books occupy 1/3 of the classical production that are usually not taken into account during the assessment exercise (section 4).

Next to the classical production, we observed that the SH produce broader content such as: reports, translations, book's reviews, working papers, scholarly work, with a high pre-eminence of the communications at congresses (around 1/3 of the stated production). Moreover, the inventory of 197 languages known amongst the SH community indicates an undeniable and irreplaceable richness and a specificity that does not exist in hard or natural sciences. This specific linguistic aspect is never taken into account with the current methods of quantitative assessment.

The evolution of the ICT and the development *Open Access in Science* according to the recommendations of the European Framework Program *Horizon 2020* is radically changing the way of producing, structuring, and disseminating knowledge. In spite of reticence about the use and the imperfection of the bibliometric existing databases, it is highly probable that the expansion of online productions and the ease with which they can be consulted will soon lead to the renewal of the assessment exercise, with more proper 'indexes', 'repositories', or 'Big data' in institutional procedures for research assessment (section 5).

Thus instead of allowing themselves to be obliged to use the considerably biased bibliometric instruments such as the *WOS*, *Scopus*, *ERIH* and national checklists as *AERES* based on the analysis of journals, it would be in the interest of researchers in the SH to continue to develop appropriated tools from databases or sources that are more open to the diversity of their publication practices. As it can be already observed the advantages of altmetrics in Science move forward the discipline's frontiers in favor of the interdisciplinarity.

8. Recommendations

In this section I precise under what conditions the quantitative assessment can be applied in the SH and express some recommendations in the light of the results of this study.

On the issue of quantitative assessment in the SH the report highlights different aspects that could be considered as basic elements for formulating recommendations to the ERC. The exercise is quite complex, because we advanced on unstable territory due to the continuous change of the modern technologies. At the same time, I have been legitimately asked to provide a clear vision of how the ERC could estimate the added-value of submitted or finalized projects in the SH.

Monitoring and assessing the research activities means finding common quality standards. One is undeniably the peer review. From the quantitative angle, we should avoid the trend of the conformity in the manner we understand the meaning of the scientific progress in SH, in terms of production and dissemination. We should not kill creativity and innovation by giving a rigid panel of assessment modeled on what was happening historically in the natural and medical sciences. In that case, classical metrics such as the calculation of the impact by the number of citations in journals' articles is simply neither suitable nor reliable (section 3). We should include in our appreciation, the inherent diversity of the disciplines and sub-disciplines and the specificities of the scientific methods of discovery within the SH (section 4). These new practices are very much connected to the revolution of the modern technologies of information and communication (ICT) (section 5).

8.1 A database dedicated to the SH

In section 4 I have referred several times to the statistical analyses produced by *RIBAC* (France). *RIBAC* is a specific observatory devoted to the activities and the production of SH' researchers at the CNRS. I experimented *RIBAC*, and suggested different proposals several times since the beginning of its development in 2009 and I consider it a good basis of what could be invented as a complement of peer review for the SH.

Another example of a similar database is the *Lattes Platform* in Brazil. *Lattes Platform* is an information system (integrated data-base, web-based query interface, etc.) maintained by the Brazilian Government to manage information on science, technology, and innovation related to

individual researchers and institutions working in Brazil. It is named after a Brazilian physicist, Cesar Lattes, and it is maintained by the federal bureau responsible for funding science and technology at the federal level, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, National Counsel of Scientific and Technological Development). Since all researchers and institutions are required to maintain their records up to date, the *Lattes Platform* can be used not only to obtain information on individual researchers but also to conduct performance evaluations at the organizational level.

Actually **such a database under control of the scholarly community does not exist at an international level**. It is deeply regrettable because those tools respect fundamental scientific standards guaranteeing the quality and the reliability of the data that produce altmetrics.

In order to adopt high-quality standards to the quantitative assessment, an online database dedicated to the SH is needed. It should objectively reflect and quantify the entire activity of the researcher in all its richness and variety including the specificities of each domain of research. The choice of the necessary data would have to be discussed among a working group of dedicated experts for Monitoring and Quantitative assessment in the SH. The data would be based on the statement of the ERC candidates when submitting a project, then at the mid-period and, again, at the final term of the project.

8.2 An Open Access Repository

Placing their research in the OA network will also cause the emergence of new forms of scientific evaluation that are better harmonized, and of which bibliometrics are only one aspect (section 5).

For these reasons the above recommended SH database could also be combined with an Open Access Repository such as the European project *OpenAIRE2020*³⁸. Indeed, the different OA

³⁸ Started in January 2015, OpenAIRE2020 will assist in **monitoring H2020 research outputs** and will be a key infrastructure for **reporting H2020's scientific publications** as it will be loosely coupled to the EC's IT backend systems. The **EC's Research Data Pilot will be supported** through European-wide outreach for **best research data management practices** and [Zenodo](#), which will provide long-tail data storage. Other activities include: **collaboration with national funders** to reinforce the infrastructure's research analytic services; an **APC Gold OA pilot for FP7 publications** with collaboration from [LIBER](#); **novel methods of review and scientific publishing** with the involvement of [hypotheses.org](#); **a study and a pilot on scientific indicators** related to open access with [CWTS](#)'s assistance; legal studies to investigate **data privacy issues** relevant to the Open Data Pilot; **international alignment**

models offer a real opportunity for these disciplines to promote their various contributions, by improving the accessibility, impact and relevance of research in SH. From now on, technical means will make distribution possible, along with the sharing of scientific work at a low cost (compared to the huge fees that are required by the commercial providers), and greater access to various resources with, as a result, an increase in the pool of citations and the production of alternative metrics. Semantic tools could also be adjusted to these metrics in order to analyse quantitatively the meaning of the indicators.

8.3 The final aim: a big data repository associated to the peer-review

The final aim would then be to build a big data Repository divided into three stages:

- ✓ Development and implementation of a specific database dedicated to SH projects (section 8.1)
- ✓ Collaboration with the European online repository of datasets harvested according to the OA protocol: **OpenAIRE2020** (section 8.2)
- ✓ Implementation of semantic tools to analyse the whole datasets (could be adjusted to 8.1 or/and 8.2)

This application will help to identify excellence by providing very rich and accurate information. It will permit to visualize dynamically the originality and the cross-disciplinary aspects of the submitted projects at the different steps of the process, and to make visible the scientific progress/impact as well as emerging fields.

To this repository of statistical and semantic information devoted to the quantitative assessment could be joined simultaneously a directory based on peer opinion, in order to ensure the certification of high-quality assessment of performance.

with related networks elsewhere with the involvement of [COAR](https://www.openaire.eu/news-events/openaire2020-press-release).
<https://www.openaire.eu/news-events/openaire2020-press-release>

8.4 The advantages of such a tool for the ERC

- ✓ to collect enriched-data as close as possible to the reality, including all the different types, from the submission, at the starting point of a project and, at the mid-term report, until the final report
- ✓ to ensure a scientific control of the whole process of collecting and retrieving the data
- ✓ to base the evaluation of the risk and the potential added-value of a submitted project on high-quality data and adapted metrics
- ✓ to give transparency to the sources and the data involved in the quantitative assessment process
- ✓ to build adapted metrics, locally relevant or depending on the specificity of each discipline
- ✓ to measure and compare the activities within the same discipline or sub-discipline at national, European, and international levels...
- ✓ to identify new trends or emerging domains of research not necessarily visible
- ✓ to obtain refined metrics from the production at the different stages of the process
- ✓ to make it possible to adjust the metrics to the evolution of science (e.g. interdisciplinarity)
- ✓ to visualize the collaborations, their evolution and the objective international impact
- ✓ to identify excellence
- ✓ to accompany the peer-review in decision-making

Bibliography

COURTIAL J.-P., 2003, "L'Association pour la mesure des sciences et des techniques (Adest) et l'évaluation de la recherche en France", *La revue pour l'histoire du CNRS*, n°9. <http://histoire-cnrs.revues.org/564>

DASSA M., 2012, "Résultats de l'enquête RIBAC 2010", *La Lettre de l'InSHS*, n°17, 3□5.

DASSA M., 2013, "Les revues scientifiques. Résultats de l'enquête Ribac 2011", *Lettre de l'InSHS*, n°22, 3□7.

DASSA M., 2014, "Les publications et interventions aux colloques des chercheurs CNRS de l'Institut des sciences humaines et sociales (InSHS)", *La Lettre de l'INSHS*, n°27, 3□9.

DASSA M., KOSMOPOULOS C., 2009, "JournalBase - Une étude comparative internationale des bases de données des revues scientifiques en sciences humaines et sociales (SHS)", *Cybergeo : European Journal of Geography*. <http://cybergeo.revues.org/22492>

DASSA M., KOSMOPOULOS C., PUMAIN D., 2010, "JournalBase - A Comparative International Study of Scientific Journal Databases in the Social Sciences and the Humanities (SSH)", *Cybergeo: European Journal of Geography*, n°484. <http://www.cybergeo.eu/index22862.html?>

GARFIELD E., 1963a, "Citation Indexes - Closing the information gap between the pure and applied sciences", *American Society for Engineering Education - Annual Meeting, June 17-21. number 103-.2, n°Unpublished. No:66.* <http://garfield.library.upenn.edu/papers/ciclosinginfogap1963.pdf>

GARFIELD E., 1963b, "Science Citation Index", *Science Citation Index 1961*, n°1, V□XVI.

GARFIELD E., 1963c, "New Factors in the Evaluation of Scientific Literature Through Citation Indexing", *American Documentation*, vol.14, n°3, 195□201.

GARFIELD E., 1964, "« Science Citation Index » A New Dimension in Indexing", *Science*, vol.144, n°3619, 649□654.

GARFIELD E., 1998, "From citation indexes to informetrics: is the Tail Now Waggin the Dog?", *Libri*, vol.48, 67□80.

GARFIELD E., 2009, "From information retrieval to scientometrics - is the dog still waggin its tail?", *Fifth International conference on WIS & Thenth COLLNET Meeting, September 13-16, 2009, Dalian, Chian*. <http://garfield.library.upenn.edu/papers/dalianchina2009.html>

GINGRAS Y., 2014, Les dérives de l'évaluation de la recherche - Du bon usage de la bibliométrie.

HARNAD S., BRODY T., 2004, "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals", *D-Lib Magazine*, vol.10, n°6. <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

HARNAD S., CARR L., GINGRAS Y., 2008, "Maximizing Research Progress Through Open Access Mandates and Metrics" (H. Kuramoto, Ed.), *Liinc em Revista*. <http://eprints.soton.ac.uk/266617/>

JACSO P., 2005, "Google Scholar and The Scientist", <http://www2.hawaii.edu/~jacso/extra/gs/>

JEANNIN P., 2003, *Revue métrique de la recherche en sciences humaines et sociales - Rapport synthétique et final de mission (1999-2003)*. France, Direction de la Recherche -Ministère délégué à la recherche et aux nouvelles technologies. http://www.revues.org/cost/images/1/12/JEANNIN_-_Rapport_final2003a.pdf

KOSMOPOULOS C., 2009, "L'Utilisation de Publish or Perish dans l'évaluation des chercheurs en SHS", *Documentaliste-Sciences de l'Information*, vol.46, n°4, 58-59.

KOSMOPOULOS C., DASSA M., 2013, "A Comparative Study of Scientific Journal Databases in the Social Sciences and the Humanities (WOS, Scopus, ERIH, AERES)" - *Presentation at the Institut National des Etudes Démographiques*, Paris, France in October 2013.

KOSMOPOULOS C., PUMAIN D., 2007, "Citation, Citation, Citation: Bibliometrics, the Web and the Social Sciences and Humanities", *Cybergeo European Journal of Geography*, n°411, 13.

LALOË F., MOSSERI R., 2009, "L'évaluation des chercheurs ! même pas juste...même pas fausse §", *Reflets de la physique*, n°13, 23-24.

PONTILLE D., TORNY D., 2013, "La manufacture de l'évaluation scientifique : algorithmes, jeux de données, outils bibliométriques", *Réseaux*, vol.1, n°177, 25-61.

SUBER P., 2004, "Open Access Overview (definition, introduction)", <http://legacy.earlham.edu/~peters/fos/overview.htm>

Annexes

Annex 1: Description of the current metrics of *Publish or Perish*

Hirsch's h-index

Proposed by J.E. Hirsch in his paper **An index to quantify an individual's scientific research output**, [arXiv:physics/0508025](https://arxiv.org/abs/physics/0508025) v5 29 Sep 2005. It aims to provide a robust single-number metric of an academic's impact, combining quality with quantity.

Egghe's g-index

Proposed by Leo Egghe in his paper **Theory and practice of the g-index**, *Scientometrics*, Vol. 69, No 1 (2006), pp. 131-152. It aims to improve on the h-index by giving more weight to highly-cited articles.

Zhang's e-index

Publish or Perish also calculates the e-index as proposed by Chun-Ting Zhang in his paper **The e-index, complementing the h-index for excess citations**, *PLoS ONE*, Vol 5, Issue 5 (May 2009), e5429. The e-index is the (square root) of the surplus of citations in the h-set beyond h^2 , i.e., beyond the theoretical minimum required to obtain a h-index of 'h'. The aim of the e-index is to differentiate between scientists with similar h-indices but different citation patterns.

Contemporary h-index

Proposed by Antonis Sidiropoulos, Dimitrios Katsaros, and Yannis Manolopoulos in their paper **Generalized h-index for disclosing latent facts in citation networks**, [arXiv:cs.DL/0607066](https://arxiv.org/abs/cs/0607066) v1 13 Jul 2006. It aims to improve on the h-index by giving more weight to recent articles, thus rewarding academics who maintain a steady level of activity.

Age-weighted citation rate (AWCR) and AW-index

The AWCR measures the average number of citations to an entire body of work, adjusted for the age of each individual paper. It was inspired by Bihui Jin's note **The AR-index: complementing the h-index**, *ISSI Newsletter*, 2007, 3(1), p. 6. The *Publish or Perish* implementation differs from Jin's definition in that we sum over *all* papers instead of only the h-core papers.

Individual h-index (original)

The Individual h-index was proposed by Pablo D. Batista, Monica G. Campiteli, Osame Kinouchi, and Alexandre S. Martinez in their paper **Is it possible to compare researchers with different scientific interests?**, *Scientometrics*, Vol 68, No. 1 (2006), pp. 179-189. It divides

the standard h-index by the average number of authors in the articles that contribute to the h-index, in order to reduce the effects of co-authorship.

Individual h-index (PoP variation)

Publish or Perish also implements an alternative individual h-index called **hI,norm** that takes a different approach: instead of dividing the total h-index, it first normalizes the number of citations for each paper by dividing the number of citations by the number of authors for that paper, then calculates the h-index of the *normalized* citation counts. This approach is much more fine-grained than Batista et al.'s; we believe that it more accurately accounts for any co-authorship effects that might be present and that it is a better approximation of the per-author impact, which is what the original h-index set out to provide.

Multi-authored h-index

A further h-like index is due to Michael Schreiber and first described in his paper **To share the fame in a fair way, h_m modifies h for multi-authored manuscripts**, *New Journal of Physics*, Vol 10 (2008), 040201-1-8. Schreiber's method uses fractional paper counts instead of reduced citation counts to account for shared authorship of papers, and then determines the multi-authored h_m index based on the resulting effective rank of the papers using undiluted citation counts.

Average annual increase in the individual h-index

As of release 4.3 *Publish or Perish* also calculates the average annual increase in **hI,norm**, called **hI,annual**. This average annual increase in the individual h-index is useful for the following reasons:

- In common with the **hI,norm** index, it removes to a considerable extent any discipline-specific publication and citation patterns that otherwise distort the h-index.
- It also reduces the effect of career length and provides a fairer comparison between junior and senior researchers.

The **hI,annual** is meant as an indicator of an individual's average annual research impact, as opposed to the lifetime score that is given by the h-index or **hI,norm**.

Source: <http://www.harzing.com/pop.htm> (March 17th, 2015)

Annex 2: 27 headlines of *JournalBase*

 JOURNAL BASE		The 27 disciplines	
<u>Anthropology, Ethnology, & Folklore</u>		Literature	
Archeology		Music et Musicology	
Art et Art History		Paedagogical & Educational Research	
Demography		Philosophy	
Law		Psychology & Cognitive Sciences	
Economics		International Relations	
Classical Studies		Religion & Theology	
Women's Studies		Political Sciences	
Gender Studies		Business & Administrative Sciences	
Oriental & African Studies		Communication Science	
Geography		Environmental Sciences	
History		Sociology	
History et Philosophy of Science		Social Statistics and Informatics	
Linguistics			

Annex 3: The number of journals per bibliometric sources and per discipline

Categories	AERES	ERIH	Scopus	WOS-AHCI	WOS-SSCI	Total	Total without duplicate per category	% of duplicatess
Anthropology, Ethnology & Folklore	256	238	123	28	71	716	394	45
Archeology	1 895	418	28	45	7	2 393	1 912	20,1
Art & Art History	0	468	1 102	204	5	1 779	1 700	4,4
Business & Administrative Sciences	686	0	509	0	231	1 426	1 034	27,5
Classical Studies	1 895	251	0	33	0	2 179	1 897	12,9
Communication Sciences	0	0	157	41	136	334	234	29,9
Demography	217	0	33	0	22	272	237	12,9
Economics	686	0	513	2	276	1 477	950	35,7
Environmental Sciences	172	0	746	7	147	1 072	921	14,1
Gender Studies	0	118	46	0	0	164	135	17,7
Geography	172	0	526	29	168	895	683	23,7
History	1 895	900	94	236	60	3 185	2 004	37,1
History & Philosophy of Science	322	166	24	37	34	583	463	2,1
International Relations	0	0	158	0	56	214	171	20,1
Law	0	0	204	0	111	315	213	32,4
Linguistics	576	579	113	109	63	1 440	657	54,4
Literature	0	795	14	345	1	1 155	963	16,6
Music & Musicology	0	166	4	68	3	241	185	23,2
Oriental & African Studies	0	0	0	36	3	39	36	7,7
Paedagogical & Educational Research	28	470	331	1	130	960	663	30,9
Philosophy	322	305	48	119	36	830	387	53,4
Political Sciences	25	0	166	10	407	608	532	12,5
Psychology and Cognitive Sciences	2 392	616	1 833	542	5	5 388	3 267	39,4
Religion & Theology	364	371	16	82	7	840	397	52,7
Social Statistics and Informatics	0	0	347	6	75	428	393	8,2
Sociology	217	0	384	21	453	1 075	855	20,5
Women Studies	0	118	0	0	29	147	122	17
Total	12 120	5 979	7 519	2 001	2 536	30 155	21 405	

Source: Dassa M., Kosmopoulos C., Pumain D., 2010

Annex 4: The 49 languages retrieved from *JournalBase*³⁹

Languages of the journals listed in <i>JournalBase</i>	
Afrikaans	Japanese
Arabic	Korean
Basque	Latin
Bosnian	Latvian
Bulgarian	Lithuanian
Catalan	Macedonian
Chinese	Malay
Croatian	Multilingual
Czech	Nepali
Danish	Norwegian
Dutch-Flemish	Persian
English	Polish

³⁹ These languages are used in the 13,525 journals listed in JB in 2013. Some of the journals allow publications in several languages and some of them only state to be ‘multilingual’.

Estonian	Portuguese
Finnish	Romanian-Moldavian Moldovan
French	Russian
Gaelic-Scottish Gaelic	Servian
Galician	Slovak
German	Slovenian
Greek	Spanish
Hebrew	Swedish
Hungarian	Thai
Icelandic	Turkish
Indonesian	Unkrainian
Irish	Welsh
Italian	Western Frisian

Annex 5: Number of journals involved in the use of the indicated language

Total of journals using the language	Languages	Total of journals using the language	Languages
10920	English	19	Latin
1950	French	19	Lithuanian
1133	German	18	Arabic
875	Spanish	16	Bulgarian
662	Italien	14	Servian
287	Portuguese	13	Icelandic
218	Dutch-Flemish	12	Estonian
169	Multilingual	11	Afrikaans
133	Polish	5	Korean
89	Russian	5	Malay
79	Croatian	5	Persian
75	Hungarian	5	Unkrainian

74	Czech	4	Hebrew
62	Swedish	3	Irish
59	Romanian- Moldavian Moldovan	- 2	Western Frisian
58	Turkish	2	Galician
55	Danish	2	Welsh
52	Chinese	2	Indonesian
46	Slovenian	2	Macedonian
45	Greek	1	Basque
40	Finnish	1	Bosnian
39	Japanese	1	Gaelic-Scottish Gaelic
29	Norwegian	1	Latvian
29	Slovak	1	Nepali
28	Catalan	1	Thai

Source: *JournalBase*, 2015

Annex 6: A sample of Zotero's network

Zotero People

[Search for people](#)

New Zotero people



Browse for people by discipline

Aeronautics (787)	History of Science and Medicine (2020)
African American Studies (2536)	Humanities (6103)
Agriculture (1506)	Information Science and Technology (9146)
American Studies (3979)	Journalism, Media, and Communication (3857)
Anthropology (4194)	Languages (4006)
Applied Mathematics (1438)	Law (3000)
Applied Physics (1218)	Library and Museum Studies (4213)
Archaeology (2211)	Linguistics (3584)
Architecture and Design (2580)	Literature (5328)
Area Studies (1610)	Materials Science (1490)
Art History (3748)	Mathematics (2565)
Astronomy and Astrophysics (1222)	Mechanical Engineering (1610)
Bioengineering (1585)	Medicine (7361)
Biology (6624)	Military Sciences (817)
Business (11260)	Molecular Biophysics and Biochemistry (1862)
Chemical Engineering (1120)	Music and Musicology (2666)
Chemistry (3045)	
Civil Engineering (1240)	

Community

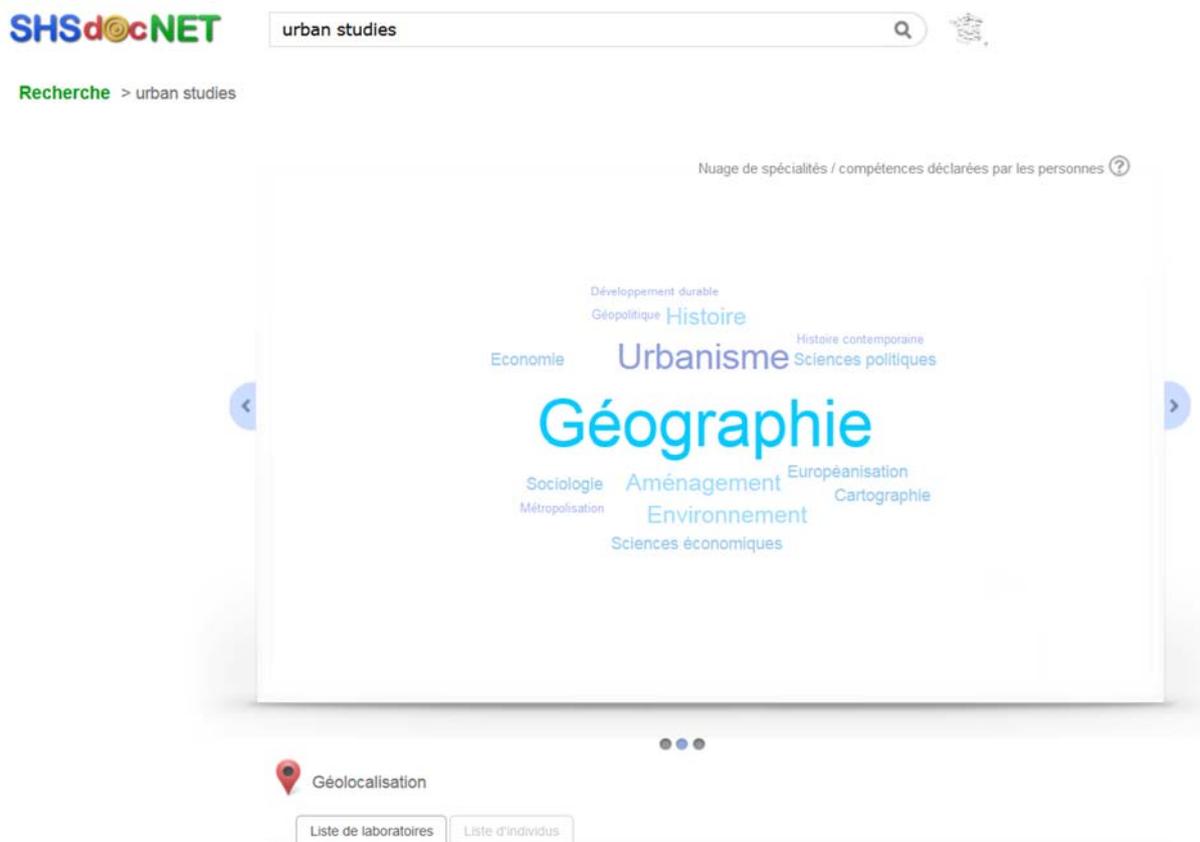
Discover researchers working on similar projects both in your field and in others.

Browse through other researchers' CVs and shared libraries for a quick view of their work and interests.

Create your own Zotero profile to help other researchers discover you and your work.

[Sign up now](#) or [log in](#)

Annex 7: A cloud of key-words generated by publications



Source: *SHSdocNet*, April 2015