Research assessment in the Humanities: problems and challenges

Research assessment is going to play a new role in the governance of universities and research institutions. Evaluation of results is evolving from a simple tool for resource allocation towards policy design.

In this respect "measuring" implies a different approach to quantitative aspects as well as to an estimation of qualitative criteria that are difficult to define. Bibliometrics became so popular, in spite of its limits, just offering a simple solution to complex problems. The theory behind it is not so robust but available results confirm this method as a reasonable trade off between costs and benefits.

Indeed there are some fields of science where quantitative indicators are very difficult to apply due to the lack of databases and data, in few words the credibility of existing information. Humanities and social sciences (HSS) need a coherent methodology to assess research outputs but current projects are not very convincing.

The possibility of creating a shared ranking of journals by the value of their contents at either institutional, national or European level is not enough as it is raising the same bias as in the hard sciences and it does not solve the problem of the various types of outputs and the different, much longer time of creation and dissemination.

The web (and web 2.0) represents a revolution in the communication of research results mainly in the HSS, and also their evaluation has to take into account this change. Furthermore, the increase of open access initiatives (green and gold road) offers a large quantity of transparent, verifiable data structured according to international standards that allow comparability beyond national limits and above all is independent from commercial agents.

The pilot scheme carried out at the university of Milan for the Faculty of Humanities demonstrated that it is possible to build quantitative, on average more robust indicators, that could provide a proxy of research production and productiivity even in the HSS.

The need for assessment

The rising costs of research and the restrictions on public budgets call for the adoption of more efficient systems of resource allocation. Indeed research assessment exercises, increasingly undertaken by national agencies and research institutions worldwide, are going to play a new role in the governance of universities and research institutions. Evaluation of results is evolving from a simple tool for resource allocation towards policy design.

There are a number of reasons for performing assessment exercises, including: evaluation of research excellence; adoption of a funding formula to distribute funds between universities and/or research institutions; ensuring accountability with regard to the use of public monies; career advancement and tenure.

In general, two approaches are currently in use for research evaluation: peer review (qualitative approach) and bibliometrics (quantitative approach). Peer review is based on the judgment of (two or more) experts that synthesizes the intrinsic quality of a scientific work, on the basis of relevance, originality, quality, or

potential socio-economic impact of research outputs. Bibliometric techniques are based on indicators, elaborated from data which can be found in *ad hoc* organised databases[1].

In this respect "measuring" scientific outputs implies a different approach to quantitative aspects as well as to an estimation of qualitative criteria that are difficult to define. It is worth noting that any attempt to introduce performance indicators of whatever type may have undesirable effects in terms of influencing which research is undertaken, the kind of outputs that are produced, the selection of the publisher or of the journal, the reference practices (IMU 2008, Arnold 2009, Abbot et al. 2010), so a strong attention is necessary when selecting bibliometric indicators, and the level of the evaluation (micro meso mega) must be taken into consideration.

Bibliometrics is a research method developed in library and information science that uses quantitative analysis and statistics in order to:

– determine the influence of single scholars or research groups, or institutions and that of single articles or groups of them

– describe the relationships between authors, publications, journals, or research fields.

Its original aim was to help libraries in their evaluation of which journals to purchase in a period of lack of resources. The journals with the most cited articles in a field became part of a core that every academic library had to have in order to satisfy the research needs of the users (Garfield 1972).

But the journals in this core also became the publications where the most and the best scientists wanted to publish to obtain visibility and reward in terms of citations.

Assuming that scholars who have to say something important do publish their findings, and that they refer in their own work to earlier work of other scholars to acknowledge intellectual debt (van Raan 2006), in research evaluation citations became a widely used measure of the *impact* of scientific publication (Cronin 1981).

So, more and more, bibliometric indexes have been used as performance indicators of institutions, research groups and individual researchers, in particular acquired importance:

– the *number of publications* produced by the scholar, (possibly divided by the scholar's academic age);

– the *number of citations* that an article received from other scholarly works, possibly divided by the number of articles published in a window (2 or 5 years).

In spite of its limits (Seglen 1997, Amin 2000, Larsen 2010, Moed 1996)[2] bibliometrics (and his main indexes Impact Factor and number of citations) became so popular just offering a simple solution to complex problems, representing a proxy of the quality of scientific works in a more efficient, transparent and comparable way. The other evaluation method, peer review, may in fact have serious shortcomings, the main of which is subjectivity of the judgment, moreover it is time consuming and expensive and above all it does not allow national or international comparison. Bibliometrics did not and does not want to be a substitute for scholarly judgment, rather it represents a tool to capture in an efficient way the judgments of a broad community. This is primarily an issue of scale. Assessing individual researchers require peer judgments too, while the same is not true for assessment of research groups or institutions at national or international level.

---

[1] Exemples of bibliometric databases are Thomson Reuter's Web of Knowledge and Elsevier Scopus among the commercial products, CiteBase http://www.citebase.org/, CiteEC, http://ideas.repec.org/top/top.series.recurse.html among free products.

[2] The most cited limits are: the window (2 years) taken into consideration is too narrow for more theoretical disciplines, the impact factor does not consider and normalize the differences in citation practices that are very different across disciplines; it does not represent a typical value of the number of citations to articles in the journal when the citation distribution is asymmetric (MacRoberts and Mac Roberts 2010), it counts citations without weighting them with the prestige of the citing journals.

Studies have demonstrated that there is a good correlation between results obtained from peer review and those possible with a bibliometric approach[3] (Abramo et al. 2009)

The distinctiveness of humanities research

Research outputs in the HSS include articles published in international and national journals (only a small part of the production) (RIA 2009), academic book chapters, books, books aimed at a much wider popular audience and non published outputs such as archeological excavations, exhibitions etc..

Researchers work generally alone or in small groups, and the *lingua franca* is normally the national language with a moderate use of English (mainly in some fields such as linguistics or logic). Much humanities research deals with specific regional or national cultural phenomena in non Anglophone journals or in non-scholarly press. The outputs are unlikely in digital format and there is a data deficit in tracking them at national level that does not allow reliable international comparisons. (Archambault et al. 2006, Nederhof 2006). Moreover, publications and referencing characteristics vary widely across disciplines.

Data for citation analysis are normally obtained from commercial databases (WOK, Scopus) or from free databases (Citebase, CiteSeer, RePEc, or Google Scholar[4]). But commercial databases have bias in favour of publications in English, have poor coverage for languages other than English, moreover the citations window (2 or 5 years for SCI) is too short for disciplines where higher citation rate is for older literature (Glanzel, Schoepflin 1999) and the main references are to non journals publications (not included in bibliometric databases).

At current time bibliometric indexes used in research evaluation in STM literature cannot be applied to the humanities. Adequate coverage of the full range of publications and agreements on how quality should be assessed would be necessary.

Bibliographic vs. bibliometric databases: data must be reliable, transparent, exhaustive, comparable

To assess excellence at European and at national level comparability is needed through reliable indicators and agreements on what is scientific literature in the Humanities. To allow comparability a bibliometric infrastructure at European level would be necessary. This could harvest the data from national bibliographic systems built according to standard criteria.

The need for a database to capture the full range of research outputs from the HSS (at European level) is well documented by a recent report commissioned by some European scientific societies (Toward a bibliographic database…2010). It aims at proposing a way forward in establishing a bibliometric database for the social sciences and humanities that should bring together in a comparable format data on number, kind and quality of the outputs in the HSS but even data about the impact on the scientific community and on the society. The first problem in the creation of such a database is to define which kind of outputs should be included. International oriented literature (outputs) is not a problem, but there is a need of shared criteria to determine the scholarly quality of national oriented literature.

There are some bibliographic databases in use for the main disciplines in the humanities, but their aim is rather literature retrieval and the metadata are collected with this primary scope. Even if some of them

---

[3] The results of the Italian First Assessment Exercise (VTR2001-2003) were compared with a bibliometric analysis developed by the authors.

[4] Citebase, CiteSeer, PubMed and CitEC are subject databases, they are quite reliable but they refer to a single discipline, Google Scholar is interdisciplinary, but is not reliable due to the uncertainty of its sources (Falagas et al. 2008, Jacso 2005, 2008, Bakklabasi et al. 2006).

record author affiliations or cited references[5], they are not projected for research assessment, they show different quality in term of data gathering and a lack of standardization and authority files. Even journal lists compiled at national or international level for evaluation exercises[6], present some problems with accuracy and reliability of content.

A bibliometric database for the HSS should be created, and it should include details about cited references and author details for all authors (not only for the first author). It needs to include a much wider range of outputs (not only journal articles, not only international outputs) and need a standardization in metadata gathering.

To assure comparability a supra-national European Organization should establish guidelines about which data are required, in what form, for which type of research output, in analogy with the "Frascati Manual" for R&D. Articles in peer reviewed (national or international) journals and books subject to peer review process could initially be included in such a database..

The experience of the Driver project[7] "whose primary objective is to create a cohesive, robust and flexible, pan-European infrastructure for digital repositories, offering sophisticated services and functionalities for researchers, administrators and the general public" could be improved and expanded. Driver is a portal that assures access to 249 open access repositories from 33 countries; all repositories are compliant with the Driver best practices for interoperability.

Institutional repositories and their role as possible collectors of bibliographic and bibliometric data

Institutional repositories (IR)[8] are digital collections of the research outputs of an institution. They collect bibliographic and administrative data about the scientific production of an institution according to international standards[9] and use a protocol that allows interoperability and comparability at national and international level (OAI-PMH)[10]. Their data are harvested from the main search engines (Google, Yahoo) but also from specialized search engines (Scientific Commons, BASE, Google Scholar). Their aim was originally to offer an overview of the whole scientific production of an institution in a single point. They can also contain the full-text of research outputs, ensuring great visibility to researchers and research groups. More than 1300 institutions in the world have an IR, even if their policies, coverage and contents are very different[11].

Recently has increased attention for IRs as sources of administrative and quantitative data whose processing could give a more comprehensive overview of the research activity of an institution. In countries where the IRs system has been developed at national level (Australia, Great Britain), they have been used as data sources for research assessment (Day 2005, Harnad 2009, ERA 2010).

---

[5] Sociological abstracts, Historical Abstracts or ECONLIT

[6] ERIH (for Europe) or ERA HCA (for Australia)

[7] http://www.driver-repository.eu/

[8] See the definition of Clifford Lynch,(2003)

[9] For example Dublin Core http://dublincore.org/

[10] http://www.openarchives.org/OAI/openarchivesprotocol.htm. OAI -PMH is a standard It allows service providers to harvest metadata from data provider and to create advanced services

[11] A recent study made an inventory of Digital repositories in Europe (van der Graaf 2009) and outlined the need of more standardization in metadata and content selection.

The University of Milan established an IR (AIR[12]) in 2005 using DSpace[13]. Since 2006 a system for the extraction and elaboration of statistics on researchers' production and productivity[14] has been associated with the IR. In 2009 a Senate statement settled mandatory self-archiving of bibliographic metadata (and full-text where possible) for every scientific work produced by researchers of the University not later than one month after publication.

AIR contains 67500 items for the window 2004-2010. Metadata values are controlled and corrected before publication and are sent every day to the national research platform[15].

Document types are articles, proceedings, books, article in books, working paper, patent, pre-print. Also exhibitions, or excavations and esteems are archived. Every item contains bibliographic data, data about the author and his affiliation (such as department and discipline), abstract, keywords and information about the quality of the outputs (journal diffusion, peer review, IF etc.)

The pilot carried out for the Faculty of Letters and Philosophy, made possible by the collaboration of the researchers who self-archived their outputs regularly since 2004, was the first in Italy, due to the lack of comprehensive data for research outputs[16] in other institutions.

The window analyzed was 2006-2008, a period in which the Faculty had 312 researchers representing the main disciplines of the Humanities. Scope of the pilot was to have an overview of the production and productivity of the Faculty for the considered period, and possibly to compare it with the results of other international experiences (UNIPUB 2009).

Outputs for professor/researcher 2006-2008								
	Total number of publications				Mean for researcher			
	Articles	Articles in book	Proceedings	Books	Articles	Articles in book	Proceedings	Books
Professors	442	863	278	358	2,25	4,40	1,4	1,82
Researchers	273	370	104	149	2,57	3,49	0,98	1,4
					mean prof. res./year 3,03			

Articles in books are the main outputs, followed by articles and books. Our results confirm what observed in other nations (Spain).

The analysis of the languages in use for these disciplines gave interesting results. Foreign languages are used very seldom, French and Spanish are as diffused as English. Only Linguistics and Logic use generally

---

[12] http://air.unimi.it

[13] www.dspace.org is an open source software developed by MIT and Hewlett-Packard that enables the creation of digital archives

[14] SurPLUS http://www.cilea.it/index.php?id=820

[15] https://loginmiur.cineca.it/

[16] The National Research site cannot guarantee a comprehensive coverage of the production of the researchers, but the main shortage is that the auto-archived data are not standardized and not controlled.

English. The researchers of the Faculty choose normally national publishers (commercial publishers, scientific societies or university presses). Only 12% of the works (205 out of 1592) are published by foreign publishers (equally distributed among English, Spanish, French or German). International orientation seems not to be very meaningful or important for the researchers of the Faculty. Research in the HSS seems to be more national oriented and devoted to a much wider public than the sole disciplinary community.

Languages other than Italian				
	Articles	Articles in book	Proceedings	Books
Professor	17%	7,10%	11,20%	8,10%
Researcher	16%	10,80%	20%	8,60%

As international studies have demonstrated (Nederhof 2006) journals in these disciplines seldom have IF.

	Total number of articles (2006-2008)	Number of articles with IF	Number of articles without IF
Professor	442	4	438
Researcher	273	9	264

Even co-authorship and collaboration degree was measured, and the results confirmed that researchers in the humanities work mainly alone or in very small groups.

Authors for output type and year			
	2006	2007	2008
Articles	1,182	1,233	1,152
Articles in books	1,542	1,194	1,688
Proceedings	1,2	1,6	1,2

The analysis of the departments productivity gave a relative uniformity as result.

Productivity 2007-2008			
	Year	Nr. of publication	Mean for researcher
Modern Philology	2007	113	3,32

	2008	113	3,32
Philosophy	2007	127	2,64
	2008	127	2,64
Geography	2007	54	4,15
	2008	52	4
Foreign Languages and Literatures (Spanish, French)	2007	134	2,73
	2008	147	3
History	2007	177	3,21
	2008	175	3,18
Classic Philology	2007	171	3,35
	2008	181	3,54
Arts and Musics	2007	92	3,83
	2008	94	3,91
Foreign Languages and Literatures (German, Slavic languages)	2007	83	3,95
	2008	85	4,04

Some data could not be taken into consideration because of its unreliability. In particular the target (national or international distribution) and the presence of peer review proved to be too subjective and not demonstrable, due to the different practices of peer review in the Humanities that brought about very different interpretations.

In Italy a weighted list of journals or an authority file of the most common publication channels does not exist as in other national systems (Norway or Australia),. Lists at institutional level seem to be quite useless because of their strictly local application. So compilation of guidelines at supranational level that give criteria to define the quality of national oriented outputs (that is the main production in those disciplines) would be very welcomed.

Scope of the pilot was not yet a qualitative evaluation of the scientific production of the Faculty, but to establish how reliable and exhaustive are data from the IR for a quantitative analysis. Future developments could include: weighted evaluation of publication channels, of esteems and of non-publication outputs. Comparability can be assured only if this non-conventional use of IR will be sufficiently spread at national and at international level.

IRs as collectors of full-text of research outputs: toward a new bibliometrics free from commercial influence?

Research impact is difficult to define (Bollen van de Sompel 2009)[17] above all in those disciplines where visibility is reached very slowly. In the Humanities books are cited more frequently than articles (even in articles) (Nederhof 2006), and data gathering about citations is very difficult in a field where paper outputs are still very common. So, the presence of full-text in IRs could be a great opportunity for the development of new techniques and methods for data gathering and measure of research results (JISC 2005).

Journal IF is not a measure for quality , rather an index of popularity (Bollen et al. 2009) and does not take into account the development, in scientific communities, of new communication channels[18] that are not considered in commercial bibliometric databases (JCR, Scopus) (Nederhof, van Leeuwen, van Raan 2009, Larsen, von Inns 2010).

Assessment in the humanities should be metric informed, not metric driven, so, more efficient, objective and transparent criteria reflecting the distinctiveness of the different communities should be used together with peer judgments, to reduce the shortcomings of peer review. To have an overview of the real value of a research work more metrics are necessary: user generated metrics (downloads, views, abstract views etc.), authors generated metrics (citation analysis, informal citations as thanksgiving, link etc.). The presence of a critical mass of open access literature allows data gathering for thesenew web metrics independent of commercial subjects, as well as increases citation advantage (Swan 2010).

Tools for measuring citations in an IR has been developed in the most recent years for single disciplines (Citebase[19] for physics, CitEc[20] for economics)

Self-archiving of full-texts in IRs offers manifold advantage especially for Humanities. Open Access aims to remove restrictions that exist on the access to articles and knowledge to the world-wide scholarly community, so that dissemination happens inmediately after publication. This art of dissemination isunthinkable in the paper world where access is limited by the short life-cycle of scientific books, but also in the digital world where copyright restrictions prevent scientific literature from a wide diffusion. A work that can be retrieved by Google or Google Scholar has more possibilities to be cited than an article in a paper journal or in an electronic journal accessible only for subscribers.

Not all publishers allow self-archiving of the final version (published version) of a work, but IRs can host more versions of the same work with various possibilities of access. That is a great advantage for national assessment exercises because IRs represent a single point of access (accessible only for evaluators) to the entire production of an institution (ERA 2010) and let free access for all to abstracts, bibliography and to the previous versions[21].

The advocacy done by the Faculty with regard of full-text self-archiving gave good results but it is soon to draw significant conclusions. The present version of DSpace does not offer statistic with the necessary granularity for bibliometric analysis. The next upgrading of the software will hopefully solve this problem.

---

[17] "Science is a gift-based economy: value is defined as the degree to which one's ideas have contributed to knowledge and impacted the thinking of others…"
[18] Those commonly described as Web 2.0

[19] http://www.citebase.org/
[20] http://ideas.repec.org/top/top.series.recurse.html .
[21] Author's version for exemple, with all suggestions of reviewers included.

Self-archiving is seen as a time consuming burden by researchers, so they are not very well disposed to this task, or they are not very accurate in the compilation of metadata. In absence of a revision, metadata from IRs are unreliable and not usable for research assessment. Although almost every university has an IR in Italy, there is no national mandatory policy for self-archiving of research outputs[22] and this can limit the possibilities of development of new metrics of the web.

Conclusions

The pilot carried out at the university of Milan gave good results regarding the opportunity to use the institutional archive as data source for internal assessment exercises. Collected data proved to be relatively reliable and exhaustive, but they will be useless for a comparison until IRs will not be present in all institutions with a strict policy in metadata and full-text gathering. IRs could represent the data source for a robust, reliable, exhaustive National research System that currently in Italy does not exist.

Abbott A., Cyranoski D., Jones N., Maher B., Schiermeier Q., Van Noorden R. (2010) Do metrics matter? Nature 465:860-862.

Abramo G., D'Angelo C.A., Caprasecca A., (2009) Allocative efficiency in public research funding: Can bibliometrics help? Research policy 38(1) 10.1016/j.physletb.2003.10.071

Amin M., Mabe M. (2000), Impact factors: use and abuse. *Perspectives in publishing* 1:1–6

Archambault E., Vignola-Gagne E., Cote G., Lavriere Gingras Y. (2006) Benchmarking scientific output in the Social Sciences and Humanities: the limits of existing databases, Scientometrics. 68(3) 329-342

Arnold Douglas N (2009) Integrity under attack, SIAM News, 42(10)

Bakkalbasi N. , Bauer K. , Glover J., Wang L. (2006), Three options for citation tracking: Google Scholar, Scopus and Web of Science, Biomedical Digital Libraries, 3:7

Bollen J, Van de Sompel H, Hagberg A, Chute R (2009) A Principal Component Analysis of 39 Scientific Impact Measures. PLoS ONE 4(6): e6022. doi:10.1371/journal.pone.0006022

Cronin B. (1981), The need for a theory of citation. *Journal of Documentation*, 37:16–24,

Day M.(2005), Institutional repositories and research assessment DOI:*10.1.1.108.2371*

ERA (2010) Submission Guidelines http://www.arc.gov.au/pdf/ERA2010_sub_guide.pdf

Falagas M. E., Pitsouni E. I., Malietzis G.A., Pappas G. (2008), Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses The FASEB Journal, 22:338-342.

Garfield E. (1972) Citation analysis as a tool in journal evaluation, Science 478: 471-479
Glanzel W., Schoepflin U., (1999) A bibliometric study of reference literature in the sciences and social sciences, Information Processing and Management, 35(1): 31-44

Harnad S.,(2009) Open Access Scientometrics and the UK Research Assessment Exercise, Scientometrics, 79,1 DOI: 10.1007/s11192-009-0409-z

---

[22] This policy is common for other national or European research funder such as ERC, Wellcome Trust, etc.

IMU (2008) Citation statistics, http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf

Jacso P. (2005), As we may search – Comparison of major features of the *Web of Science, Scopus, and Google Scholar* citation-based and citation-enhanced databases, Current Science 89: 1537-1547

Jacso P., (2008), Google Scholar revisited, Online information Review, 32(1): 102-114

JISC Scholarly Communication Group (2005) Open Access Citation Information. Final report – Extended version
http://www.jisc.ac.uk/uploaded_documents/iv)%20OA%20Citation%20Information%20FINAL%20Extended%20Report.DOC

Larsen P.O, von Inns M. (2010), The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index, Scientometrics DOI 10.1007/s11192-010-0202-z

MacRoberts M.H. and MacRoberts B.R. (2010) Problems of citation analysis: a study of uncited and seldom cited influences JASIST 61(1): 1-13

Moed, H. F., & Van Leeuwen, Th. N. (1996). Impact factors can mislead. *Nature, 381*, 186.

Nederhof A.J. (2006) Bibliometric monitoring of research performance in the social sciences and the Humanities: a review, Scientometrics, 66, 1(2006) pp. 81-100

Nederhof A. J., van Leeuwen T. N., van Raan A. F. J. (2009) Highly cited non-journal publications in political science, economics and psychology: a first exploration, Scientometrics DOI 10.1007/s11192-009-0086-y

Reale E., De Filippo D.A., Gomez I., Lepori B., Probst C., Potì B., Primeri E., Sanz Casado E.(2009) UNIPUB projectMethodologies for the characterisation of the publication output of higher education institutions using institutional databases Final report, PRIME
RIA – Royal Irish Academy (2009) Developing key performance indicators for the humanities, http://www.ria.ie/getmedia/9fdc0fbf-3a0a-4185-975b-98b61dcb54a5/humanities-for-web.pdf.aspx

Seglen, P: O (1997), Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314:498–502

Swan A. (2010) The Open Access citation advantage: studies and results to date (technical report) http://eprints.ecs.soton.ac.uk/18516/

Toward a bibliometric database for the social sciences and humanities – a European scoping project (2010)
Van der Graaf, Maurits(2009) *The European Research Repository Landscape 2008: Inventory of Digital Repositories for Research Output*. Amsterdam University Press http://dare.uva.nl/aup/nl/record/316871

van Raan A.F.J (2006) Measuring science. Capita selecta of current main issues. In. Moed H.F., Glanzel W, and Schmoch U., editors, *Handbook of quantitative science and technology research* pages 19–50. Kluwer