RESSH2017

2nd International Conference on Research Evaluation in the Social Sciences and Humanities

University of Antwerp, Campus Middelheim, Antwerp, Belgium
6 - 7 July 2017

EVALHUM
Working for the SSH

ECOOM
Expertisecentrum O&O Monitoring

Universiteit Antwerpen

RESSH2017

The EvalHum Initiative is pleased to welcome you at its second international conference on Research Evaluation in the Social Sciences and the Humanities (RESSH), to be held at the University of Antwerp, Belgium, on the 6th and 7th of July 2017.

This conference follows the highly successful first conference that was held in Rennes, France, in June 2015. The aim of this event is to bring together a wide range of researchers and stakeholders interested in questions of research evaluation and the societal impact of the social sciences and the humanities (SSH).

Conference themes

The development of research evaluation protocols and performance-based funding systems is rooted in policy decisions aiming to improve research and the research system as a whole. However, the implementation of such policies has attracted a great deal of criticism, especially from the SSH. The purpose of the RESSH conference is to explore whether evaluation can be useful to SSH researchers, whether it seeks to improve the quality and relevance of research or whether it is simply a management tool for the allocation of funds or imposition of policy.

Topics of interest include (but are not limited to):

- Evaluation of societal relevance of SSH research
- Notions of quality in SSH research
- Diversity in SSH research and its relation to assessment
- Bibliometric indicators for the SSH
- Alternative metrics of SSH research impact
- Data sources on SSH research output
- Evaluation practices in SSH
- Role of evaluation in knowledge production processes
- Effects of evaluation and performance-based funding on SSH research and researchers
- Peer review in the SSH
- Open Access for SSH and its potential for assessment of research
- The international dimension in SSH research
- Characteristics of dissemination channels

Conference Organisers

EvalHum is an open initiative, aiming at promoting the study of SSH research and increasing its visibility among scholars of other fields and disciplines, the lay public and stakeholders and decision makers. Evaluation appears as an interesting tool to achieve this. EvalHum seeks to break both locked-in and locked-out syndromes affecting the SSH, through bringing together all actors, with an emphasis on research evaluators, but also on the shop floor of the SSH researchers. http://www.evalhum.eu/

ECOOM is the Flemish Centre for Research & Development Monitoring. It is an interuniversity consortium with participation of all Flemish universities (KU Leuven, UGent, VUB, UA and UHasselt). Its mission is to develop a consistent system of R&D and Innovation (RD&I) indicators for the Flemish government. ECOOM-Antwerp is the branch of ECOOM that specializes in monitoring of Social Sciences and Humanities. https://www.ecoom.be/en

Scientific committee

- Paul BENNEWORTH
- Tim ENGELS
- Ioana GALLERON
- Elea GIMENEZ TOLEDO
- Raf GUNS
- Emanuel KULCZYCKI
- Michael OCHSNER
- Thed VAN LEEUWEN
- Linda SILE
- Jolanta SINKUNIENE
- Gunnar SIVERTSEN
- Jack SPAAPEN
- Mimi URBANC
- Frederik VERLEYSEN
- Geoffrey WILLIAMS
- Alesia ZUCCALA

Sponsors

RESSH 2017 is financially supported by

Program overview

Thursday, July 6

	G.010	G.005	G.006
8:30-9:00	Registration		
9:00-9:15	Welcome and opening		
9:15-10:15	Pre-conference workshop on book evaluation		
10:15-10:45	Coffee break		
10:45-12:00	Pre-conference workshop on book evaluation: panel discussion		
12:00–13:30	Lunch		
13:30-14:45	Keynote #1		
14:45-15:15	Coffee break		
15:15-17:20		Societal impact of SSH	Bibliometrics of SSH
19:00-21:00	Conference dinner		

Friday, July 7

	G.010	G.005	G.006
9:00-10:15	Keynote #2		
10:15-10:45	Coffee break		
10:45-12:25		Alternative metrics	Case studies per country
12:25–14:00	Lunch		
14:00-15:40		National research assessments	Peer review
15:40-16:10	Coffee break		
16:10-17:00		Databases	Evaluation in Europe
17:00-17:30	Closing		
19:00-21:00	Beer tasting		

RESSH2017

Program

Thursday, July 6

Pre-conference workshop on book evaluation
9:15 - 10:15 / 10:45 - 12:00 > G.010

Chair: Tim Engels

9:15 - 10:15 presentations

- Alesia Zuccala, Mads Breum, Kasper Bruun and Bernd T. Wunsch
  The Functional Requirements for Bibliographic Records (FRBR) Framework for Indexing Monographs: Implications for the Book Citation Index™ and Metric Evaluations
- Ioana Galleron, Geoffrey Williams, Elea Giménez-Toledo, Jorge Mañana-Rodríguez and Antonella Basso
  The language of books in the SSH: publication trends in France, Italy and Spain
- Frederik Verleysen
  How arbitrary are the weights assigned to publication types in national publication indicators? Comparative reflections on the Flemish case

10:45 - 12:00 panel discussion

- Geoffrey Williams (moderator)
- Judit Bar-Ilan
- Jon Holm
- Emanuel Kulczycki
- Hannelore Vanhaverbeke

Keynote #1 *13:30 - 14:45 > G.010*

Howard D. White
Libcitations, WorldCat, and Cultural Impact

Societal impact of Social Sciences and Humanities
15:15 - 17:20 > G.005

Chair: Jack Spaapen

- Nelius Boshoff and Mpho Sefatsa
  Creating impact through 'productive interactions': An example from South African research on maritime piracy
- Eiríkur Smári Sigurðarson
  Drowning by Numbers: Evaluating Social Capacities
- David Budtz Pedersen, Jonas Grønvad and Rolf Hvidtfeldt
  Mapping the societal impact of SSH – a literature review

- Gunnar Sivertsen
  Frameworks for understanding the societal relevance of the humanities
- Paul Benneworth, Julia Olmos-Peñuela and Reetta Muhonen
  Towards a common understanding on the societal impact of SSH research

Bibliometrics of Social Sciences and Humanities
15:15 - 17:20 > G.006

Chair: Alesia Zuccala

- Emanuel Kulczycki, Ewa Rozkosz and Aneta Drabek
  Ostensible internationalization of journals in the social sciences and humanities in Poland as a result of the Polish Journal Rankings
- Thomas Franssen and Paul Wouters
  Representing the humanities in bibliometric scholarship
- Elea Giménez-Toledo, Jorge Mañana-Rodríguez, Elba Mauleón Azpilicueta, Daniela De Filippo and Elias Sanz-Casado
  Proposal of an assessment model for scholarly publishers applied to university scientific output
- Janne Pölönen, Tim Engels, Raf Guns, Gunnar Sivertsen and Frederik Verleysen
  SSH journal publishing in Flanders and Finland

Friday, July 7

Keynote #2 *9:00 - 10:15 > G.010*

Askold Ivantchik
The reform of the Academy of sciences and problems of research evaluation in Russia: between bibliometrics and expert evaluation

Alternative metrics
10:45 - 12:25 > G.005

Chair: Emanuel Kulczycki

- Stacy Konkiel, Nicky Agate, Christopher Long, Simone Sacchi, Jason Rhody and Rebecca Kennison
  Towards values-based evaluation in the humanities and social sciences
- Judit Bar-Ilan
  Reader and Citation Counts of Journal Articles in Linguistics
- Evangelia Lipitakis
  Usage-based indicators for the Social Sciences and Humanities in Web of Science
- Jadranka Stojanovski, Franjo Pehar, Tomislav Jagušt and Draženko Celjak
  Toward alternative metrics of SSH research impact: A comparison of visit, download and citation data

Case studies per country
10:45 - 12:25 > G.006

Chair: Gunnar Sivertsen

- Elea Giménez-Toledo and Jorge Mañana-Rodríguez
  How Spanish book publishers select their monographs?
- Jaroslav Šušol and Marta Dušková
  Impact of performance-based funding on publication patterns: Money changes everything...?
- Michael Ochsner and Miso Dokmanovic
  Quality criteria and research obstacles in the SSH in Macedonia

National research assessments
14:00 - 15:40 > G.005

Chair: Paul Benneworth

- Antonio Ferrara and Marco Malgarini
  Changing publication practices in SSH: Evidence from two consecutive national research assessment exercises (VQR 1, 2004-10; VQR 2, 2011-14)
- Marta Natalia Wroblewska
  Staging research impact. How academics write and talk about the wider impact of their research in the context of REF

- Jon Holm and Heidi Dybesland
  Research groups in SSH – A typological analysis of 220 groups submitted to national research evaluations in Norway
- Eriko Amano, Ayako Fujieda, Natsuko Inaishi, Toshiro Kamiya, Akiko Morishita, Asa Nakano, Yoshimi Osawa and Yu Sasaki
  The Quest for 'Research Quality' in the Humanities and Social Sciences: A Japanese perspective on the Research Excellence Framework (REF)

Peer review
14:00 - 15:40 > G.006

Chair: Raf Guns

- Imre Lendák and Karolina Lendak-Kabok
  Game theory, crowdsourcing and the peer review system
- Thed van Leeuwen, Thomas Franssen, Clara Calero Medina, Giovanni Colavizza, Ismael Rafols and Nicolas Robinson-Garcia
  In search of visibility. How do publication culture, collaboration patterns, language of publication and locality of topics influence research assessments in the SSH ?
- Michael Ochsner and Sven E. Hug
  Quality Criteria for Ex-Ante Evaluation of Research Proposals from Young Humanities Scholars
- Martin Reinhart
  Are theories of peer review fit for the Humanities and Social Sciences

Databases
16:10 - 17:00 > G.005

Chair: Elea Giménez-Toledo

- Linda Sīle
  Databases and repositories for Social Sciences and Humanities research output: describing the present, discussing the future
- Hanna-Mari Puuska, Janne Pölönen, Tim Engels and Gunnar Sivertsen
  Towards integration of European research information

Evaluation in Europe
16:10 - 17:00 > G.006

Chair: Ioana Galleron

- Michael Ochsner, Emanuel Kulczycki and Aldis Gedutis
  SSH research evaluation in Europe: a classification
- Marc Vanholsbeeck
  The contradictions of the European (Open) Science policies with regard to the evaluation of research and publications in the social sciences and the humanities

Keynotes

Keynote #1: Libcitations, WorldCat, and Cultural Impact

Howard D. White

Union catalogs, which are created by library cooperatives, give bibliographic information on particular books and name the member libraries that hold them. For any given book, the libraries holding it can be counted, as if they were citations to it. Hence my 2009 proposal that these holdings counts, termed "libcitations," can complement citation counts in evaluating a book's cultural impact. The proposal was specifically intended to benefit book-oriented authors in the social sciences and humanities, whose books citation measures may undervalue. Here, libcitation analysis will be illustrated with titles from OCLC's WorldCat, the world's largest union catalog. Some show its relevance to European scholarship. Others come from a special dataset in which thousands of books, all cited by humanities scholars in Scopus, can be ranked by their libcitation counts in main Dewey classes. The ranking permits titles from different points of the libcitation distribution to be qualitatively compared. The notion of "cultural impact" takes on concrete meanings in this context—meanings related to fame and fashion. At the same time, the strengths and weaknesses of libcitation analysis can be better understood, as well as misconceptions about it. WorldCat has features that permit authors to gather libcitation data on their own works, but, like the Web of Science, it also has complexities they should know about.

Keynote #2: The reform of the Academy of sciences and problems of research evaluation in Russia: between bibliometrics and expert evaluation

Askold Ivantchik

The system of the organization of researches in Russia had dramatically changed after 2013 when the reform of the Russian Academy of sciences and its research institutions has started. The new system needs developing of new methods of research evaluation; these methods are actively debated in Russia in the last three years in the research community, as well as among officials and politicians. The paper is devoted to the analysis of this debate and to the first experience of the research evaluation in Russia according to new rules, especially in the field of humanities.

Pre-conference workshop on book evaluation

The Functional Requirements for Bibliographic Records (FRBR) Framework for Indexing Monographs: Implications for the Book Citation Index™ and Metric Evaluations

Alesia Zuccala, Mads Breum, Kasper Bruun, and Bernd T. Wunsch

Introduction

In the past, bibliographic data and citation data pertaining to books were difficult to retrieve. Now, as digital resources have improved, so has the priority to advance book-related metrics. This is partly due to the introduction of Thomson Reuter's Book Citation Index™ (BKCI) (Adams & Testa, 2011) and the addition of books to Elsevier's Scopus. These commercial databases; however, are not the 'be-all and end-all' for the discerning bibliometrician. Recent assessments of the BKCI (in particular) point to numerous indexing problems, which can lead to flawed evaluations (Gorraiz et al., 2013; Leydesdorff & Felt, 2013; Torres-Salinas et al., 2014). Still, researchers continue to use the BKCI or Scopus, and work mainly with book citations from journal articles (Hammarfelt, 2011; Zuccala et al., 2014), or choose alternative resources, like Google Books (Kousha & Thelwall, 2009), Google Scholar (Kousha & Thelwall, 2011) and OCLC WorldCat (Torres-Salinas & Moed, 2009; White et al., 2009). Concerted efforts are even being made to assess data that has been retrieved from multiple resources (e.g., Zuccala & Cornacchia, 2016).

The bibliometrics community is making rapid progress, but there are still several issues that need to be addressed in order to produce more effective evaluations. One issue in particular is central to all previous studies combined: *regardless of where and how bibliographic and citation data are collected, it is essential to recognize that books often belong to bibliographic 'families'*.

Since 'bibliographic families' can be examined both theoretically and empirically, the aim of our study is to examine and explain several interrelated concepts linked to a family-oriented entity-relationship model, known as the Functional Requirements for Bibliographic Records (FRBR). Here, we have chosen to use this model to illustrate the extent to which books, as complex entities, are not always indexed accurately with appropriate metadata. In the second part of our study, we will present some data collected and assessed from the BKCI, OCLC-WorldCat, and Goodreads, and use this data to demonstrate why a robust model is necessary, first for the practice of indexing books, and more critically for the practice of book-oriented metrics. The empirical aspect of our research is based on the following question: *Do books currently indexed in the Book Citation Index have adequate metadata and/or data designed to reflect inherent familial components and relationships?*

The FRBR Model and Bibliographic Families

In the Functional Requirements for Bibliographic Records (FRBR) the term 'work' is an abstract entity, which serves as the focal point for a full conceptual model of the bibliographic universe (Tillett, 2005). The FRBR model was developed in 1998 by a study group affiliated with the International Federation of Library Associations and Institutions (IFLA). Originally this model was meant to serve as "a generalized view… independent of any cataloguing code or implementation" (Tillett, 2005, p. 24). Now it is often recommended for the restructuring of catalogs:

> *The number of records we make is a decision made up front by the cataloger based on local policies reflecting local user needs. We may choose to catalog at various levels: the collection of works (FRBR calls this an aggregation), an individual work, or a component of a work. At the description level, we may include a description of all the parts and should provide access to each component. At the component level, we should provide a link to relate to the larger 'whole.' (Tillet, 2005, p. 27)*

Figure 1 illustrates what the FRBR entity-relationship model might look like as a guide to evaluating the current structure of the BKCI. This is an adapted version of Tillett's (2001) original figure, published in *Relationships and the Organization of Knowledge*. Here the figure only represents scholarly 'work' and indicates the cut-off point when an expression is recognized as a 'new work'. Below, we also provide a list of defined concepts.

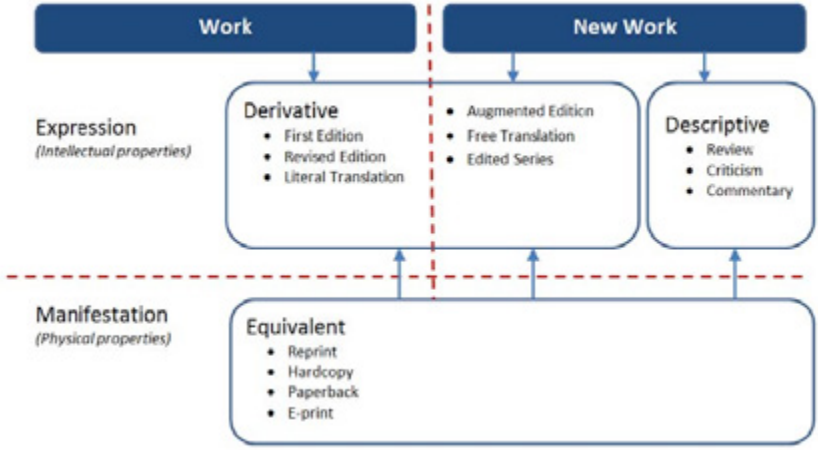


Figure 1. Modified FRBR model of 'bibliographic' families for a scholarly work.

1. **First Edition:** the emblematic or original version of a work as an intellectual contribution
2. **Revised Edition:** an edition that includes small corrections made to the original work
3. **Literal Translation:** a direct translation of the original language text into a target language text whereby the intellectual domain and the historical-temporal context of the original work is recognized and maintained (Pellizzi, 2015).

4. **Augmented Edition:** a new edition of a 'work' that is based on an earlier work with augmented or new intellectual content
5. **Free Translation:** an approach to translating a text, which intentionally recognizes the cultural gap between the "intellectual world of the author and that of the translator" (Pellizzi, 2015, p. 10); it modifies parts of the original language text, so that it appeals differently to the audience of the target language text.
6. **Edited Series:** by default every new expression of an edited series with new intellectual content will become a 'new work', even if the title of the edited series remains the same.
7. **Review:** a focused piece of work written by a new author to describe and review the intellectual content of the original, emblematic 'work' or one of its expressions (e.g., a book review)
8. **Criticism:** an extensive piece of work written by a new author which critically evaluates the intellectual content of the original, emblematic 'work' or one of its expressions in connection with other similar works (e.g., literary criticism)
9. **Commentary:** a work that explains and annotates an original 'work' (e.g., a commentary on one or more expressions of the Bible).

Research

Our research focuses primarily on the BKCI, but in order to assess its reliability as a data source for bibliometric analyses, we have chosen to compare it to three other catalogs: 1) the Danish Pure repository of scholarly research outputs, 2) the OCLC-WorldCat, and 3) Goodreads. Each database/catalog was selected for a specific reason.

The Danish Pure repository is a data repository for research outputs that was designed to support the performance-based evaluation system in Denmark. As of 2009, all Danish scholars in universities across the country have been required by law to register their scholarly publications in Pure. Each year, performance 'points' are then calculated and used to determine the amount of leftover government funding (25% of the total funding) to be distributed across departments or research centers. Monographs are amongst the document types included and each register earns a department or research center 5.00 points (level 1 authority publisher) or 8.00 points (level 2 authority publisher) towards extra funding (Giménez-Toledo, 2016). The dataset retrieved for this study was a set of monographs that had been registered in Pure between the years of 2005-2015. Our main reason for working with data from Pure was to examine its current indexing quality, and to determine the extent to which books published by Danish scholars are indexed (or have been indexed) also in the new BKCI.

The OCLC-WorldCat and Goodreads were also chosen for this study because both catalogs comply to some degree with the FRBR standard. The BKCI does not; thus by matching ISBNs and extracting all related data from these two extra databases, it possible to assess the extent to which the BKCI is an accurate index of monographs as family-based entities.

References

Adams, J., & Testa, J. (2011). Thomson Reuters Book Citation Index. In E. Noyons, P. Ngulube & J. Leta, Eds.), *The 13th Conference of the International Society for Scientometrics and Informetrics* (Vol. I, pp. 13-18). Durban, South Africa: ISSI, Leiden University and the University of Zululand.

Giménez-Toledo, E., Manana-Rodrıguez, J., Engels, T. C. E., Ingwersen, P., Polo-nen, J., Sivertsen, G., Verleysen, F.T. and Zuccala, A. A. (2016). Taking Scholarly Books into Account. Current Developments in Five European Countries. *Scientometrics, 107(2)*, 685-699.

Gorraiz, J., Purnell, P., & Glänzel, W. (2013). Opportunities and limitations of the book citation index. *Journal of the American Society for Information Science and Technology, 64(7)*, 1388–1398.

Hammarfelt, B., (2011). Interdisciplinarity and the intellectual base of literature studies: Citation analysis of highly cited monographs*. Scientometrics, 86(3)*, 705-725.

Kousha, K. & Thelwall, M. (2009). Google book citation for assessing invisible impact? *Journal of the American Society for Information Science and Technology, 60(8),* 1537-1549.

Kousha, K. & Thelwall, M. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology, 62(11)*, 2147-2164.

Pellizzi, F. (2015). Art historical and anthropological translation: Some notes and recollections. *Art in Translation, 4(1)*, 9-16.

Tillet, B. (2001). Bibliographic relationships. In C. A. Bean and R. Green (eds.) *Relationships in the Organization of Knowledge* (pp. 19-35). Dordrecht: Kluwer Academic Publishers.

Tillet, B. (2005). What is FRBR?. A conceptual model for the bibliographic universe. *The Australian Library Journal, 54(1),* 24-30. DOI:10.1080/00049670.2005.10721710.

Torres-Salinas, D., Robinson-Garcia, N., Cabezas-Clavijo, A. & Jimenez-Contreras, E. (2014). Analyzing the citation characteristics of books: edited books, book series and publisher types in the Book Citation Index. *Scientometrics, 98(3),* 2113–2127.

White, H., Boell, S.K, Yu, H., Davis, M., Wilson, C.S. and Cole, F.T.H. (2009). Libcitations: a measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology, 60(6)*, 1083-1096.

Zuccala, A. A., & Cornacchia, R. (2016). Data matching, integration, and interoperability for a metric assessment of monographs. *Scientometrics,108(1),* 465–484. DOI: 10.1007/s11192-016-1911-8

Zuccala, A., Guns, R., Cornacchia, R., & Bod, R. (2014). Can we rank scholarly book publishers? A bibliometric experiment with the field of history. *Journal of the American Society for Information Science and Technology, 66(7),* 1333-1347.

The language of books in the SSH: publication trends in France, Italy and Spain

Ioana Galleron, Geoffrey Williams, Elea Giménez-Toledo, Jorge Mañana-Rodríguez and Antonella Basso

Internationalisation has often become to mean publishing in English. However, most SSH Scholars working in romance languages tend to have a point in common, they can at least read each other's languages. This means that internationalism without English is a real possibility. The problem is to know the extent that scholars publish within languages other than English and whether that internationalism is recognised by the system.

In order to tackle this issue, three complementary approaches will be used.

This presentation will first try to quantify this form of internationalism in publications of SSH scholars, through using data from France, Italy and Spain. All SSH sectors will be scrutinized with regards to papers of any form published in 2010 and 2011. The sources of information are quite heterogeneous, since for France we will extract information from HAL, which is a repository, incomplete by its own nature, while data for Italy has been systematically collected via the universities. In Spain, data will be extracted from the annual reports published by the National Association of Universities (CRUE) and from some institutional reports (i.e. CSIC. UC3M, etc.) However, samples are of approximately the same size and can be compared, as such.

In a second part, our focus will be on books (whether monographs or collected essays), since these are known as being a major publication outlet for SSH scholars. Taking as starting point the national databases for the book industry, we propose to analyse fields related to translations and languages of books in such a way we can conclude how relevant are the cultural exchanges in our countries, how is the influence of Italian research in France or Spain (or viceversa), etc.

Last, we will look at the evaluation framework in these three countries. Is there any provision for a specific evaluation/ reward for publication in other languages than national/ English? To what extent is it possible to observe a correlation between such national legislative frameworks and publishing behaviours?

How arbitrary are the weights assigned to publication types in national publication indicators? Comparative reflections on the Flemish case

Frederik Verleysen

Performance-based research funding systems (PRFS) making use of a publication indicator usually assign a different weight to various publication types included in the national publication output. Such weight differentiation is especially important to many research communities in the social sciences and humanities (SSH), where there is far less dominance of the international journal article than is the case in scientific, technical or medical fields. Instead, chapters, monographs and edited books represent a significant share of the peer reviewed output and are usually assigned an equal or higher weight in publication indicators for the SSH than are journal articles (Giménez-Toledo et al., 2016; Verleysen, Ghesquière, & Engels, 2014).

The political decisions on the weighting of publication types in various countries have mostly been made based on consultations of SSH research communities and/or on the informed opinion of expert bibliometricians. Yet it seems that assigning a higher weight to books in publication indicators does not so much aim to reflect books' central position in the scholarly communication system in terms of output or impact –which can be hard to measure-, but is mostly defended on the basis of the extra effort it takes, on average, for scholars to conduct research for and to write a book compared to a journal article (Verleysen, 2016).

A comparison of the publication indicators of the well-documented PRFS's of Denmark, Finland, Flanders (Belgium) and Norway shows that the weight assigned to publication types, especially books, varies considerably between these countries (Giménez-Toledo et al., 2016). However, their publication patterns in SSH have been shown to be rather similar (Sivertsen, 2016). From this we conclude that there seems to be a degree of arbitrariness in assigning weights to publication types. Determining how arbitrary such weights really are is more challenging. Yet it is possible, based on Flemish VABB-SHW data, to propose a rudimentary bibliometric indicator of the effort invested in researching and writing publications across the SSH: the median page length of the five included publication types (articles, monographs, edited books, chapters and proceedings papers). In assuming that writing one page takes the same effort, irrespective of publication type or field of research, page length ratios between publication types can thus be compared to weight ratios in the Flemish funding model. Comparisons with the weighting schemes used in other countries are also possible.

For our presentation at RESHH-2017, we propose to do the following:

- to present a comparative overview of publication type weighting schemes and their histories used in various countries, with a focus on the Flemish case;
- to present median page length as a rudimentary bibliometric indicator reflecting scholarly effort in the genesis of various publication types and to compare this to the

weights assigned to publication types used in funding models;
- to critically assess the use of unified weighting schemes at various aggregation levels and across SSH fields.

In conclusion, we will reflect on possible future developments in designing and redesigning weighting schemes for publication types used in publication indicators for the SSH.

References

Giménez-Toledo, E., Manana-Rodriguez, J., Engels, T. C. E., Ingwersen, P., Pölönen, J., Sivertsen, G., Verleysen, F.T., Zuccala, A. (2016). Taking scholarly books into account. Current developments in five European countries. *Scientometrics, 107*(2), 685-699. doi:DOI: 10.1007/s11192-016-1886-5

Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics, 107*(2), 357-368.

Verleysen, F. T. (2016). *Books in the social sciences and humanities. Analyses of scholarly publication patterns based on the VABB-SHW.* (doctoral dissertation), University of Antwerp, Antwerp, Belgium.

Verleysen, F. T., Ghesquière, P., & Engels, T. C. E. (2014). The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In W. Blockmans & al. (Eds.), *The use and abuse of bibliometrics* (pp. 115-125): Academiae Europaea; Portland Press.

Societal impact of the SSH

Creating impact through 'productive interactions': An example from South African research on maritime piracy

Nelius Boshoff and Mpho Sefatsa

Introduction

The focus of this study is on the impact of research relating to the policing of maritime piracy in South African waters. The relevant research was completed in 2006 as part of a PhD by a South African academic (Professor Henri Fouché [HF]). HF had served for 24 years in law enforcement before his academic appointment and enrolment for PhD studies. His PhD highlighted a number of challenges confronting maritime security: that international law in relation to South Africa was insufficient to effectively deal with piracy in South African waters; that the capacity of South Africa's police was insufficient to deal with incidents of piracy; and that South Africa lacked a coordinated counter strategy to deal with maritime piracy. One set of PhD recommendations, 'partnershipping', called for co-operation and agreements between African states and also identified tasks for a number of role-players, including the International Police Organisation (Interpol) (Fouché, 2006).

The current study investigated the uptake and impact of HF's research and specifically the nature of his interactions with relevant parties. The SIAMPI approach to research impact assessment – which traces so-called 'productive interactions' (Spaapen & Van Drooge, 2011) – was considered a useful tool for this exercise.

The SIAMPI approach and 'productive interactions'

SIAMPI stands for the *S*ocial *I*mpact *A*ssessment *M*ethods for research and funding instruments through the study of *P*roductive *I*nteractions between science and society. The project was supported under the Seventh Framework Programme of the European Union. According to SIAMPI, for impact to occur, there needs to be some 'interaction' between a researcher and a stakeholder. The interaction can be direct (e.g. a face to face meeting or email correspondence), indirect (e.g. the reading of a publication or information received via a third-party) or financial (e.g. a research contract or economic contribution) (Spaapen & Van Drooge, 2011). An interaction is deemed 'productive' when the stakeholder makes an effort to engage with the research, based on that interaction. Should the productive interaction lead to the stakeholder doing things differently, the research is said to have an impact (Molas-Gallart & Tang, 2011). Impact in SIAMPI is thus very much different from what is commonly encountered in programme evaluation theory. Impact is not about demonstrating long-term changes and establishing strong claims of attribution. For SIAMPI, social impact "is often impossible to measure" and therefore the approach is not "fixated on societal impact itself" (http://www.siampi.eu/642.bGFuZz1FTkc.html). Instead, the objective is to learn by focusing on the processes responsible for creating impact.

Although the notion of 'productive interactions' is well referenced in the research impact literature, it is notable that, other than the case studies that were part of the original SIAMPI project (De Jong et al., 2014; Molas-Gallart & Tang, 2011), there have been no further empirical demonstrations of the SIAMPI approach (as far as could be detected in

the peer-reviewed literature). Our study addresses this vacuum and makes a contribution in two ways. Firstly, our object of assessment was a set of interactions of a single researcher (an individual) and not that of a research team or project or organisation. Secondly, instances of direct and indirect interactions, and their effects, were largely obtained through the self-documentation efforts of HF (e.g. emails and other letters, publications and CV).

Method

The study approach is best summarised as an evidence-based narrative of impact contribution, structured according to instances of direct and indirect interactions between HF and academic and non-academic stakeholders. The approach involved the following:

- HF compiled a portfolio of his research interactions and potential impacts, in the form of a 134-page file. The file included a 12-page CV, 5 emails, 5 meeting invitations/agendas/reports, 3 news clips, 9 letters of commendation/appreciation, and 10 other letters.
- The file contents were systematically coded (by the authors) in terms of four elements: direct interactions (46 examples), indirect interactions (29), enabling factors (15) and effects (18). The exercise was confined to the period 2005-2014. Sources of evidence were also recorded (see Table 1).
- An interview was conducted with HF, to clarify uncertainties, to elaborate on the documented interactions, and to enquire about other interactions not covered by the documentation provided.
- A draft 'contribution narrative' was constructed, i.e. a story of contribution using the coding elements and their linkages as anchor points.

Table 1. Example of coding elements for the year 2012.

Codes	Interactions and effects
In	HF co-authored an article on 'Investigating sea piracy: Crime scene challenges' in the *World Maritime University Journal of Maritime Affairs*. [S1]
En	School of European Studies at Cardiff University placed article on their website of publications dealing with maritime piracy (http://piracy-studies.org) [S2]
Ef01	Lecturer at Cardiff University: "I am conducting research on piracy largely from a political science perspective. And hence found your alternative (and very practical) take on the issue exceptionally useful." [S2]
Di	HF participated in United Nations expert meeting, Vienna, on the investigation and prosecution of transnational organised crime at sea. [S1, S3]
Ef02	"The outcomes of the expert meeting will provide input to a final report with recommendations that provide a better understanding of the challenges faced in tackling international organized crimes committed at sea and identify areas for possible future action. These findings and recommendations will be reported to the meeting of the Commission on Crime Prevention and Criminal Justice in April 2013." [S3]

Codes: Di = direct interaction; In = indirect interaction; En = enabling factor; Ef = effect
Sources of evidence: [S1]CV; [S2]Email letter; [S3]Meeting invitation

First results and conclusion

The study identified a number of productive interactions, two of which can be considered 'success stories' to the extent that the associated effects reflect impact as defined by SIAMPI. The first case is built around direct and indirect interactions between HF and Interpol, thereby contributing to Interpol's endorsement of a law enforcement approach to maritime piracy. *Security Watch Africa*, a media NGO based in Nigeria that covers security issues and events across Africa, futures in the second case. Here the interactions led to the maritime domain becoming part of the NGO's media coverage. In the first case, impact was made possible through direct interactions that existed long before the PhD research. In the second case, the initial direct interaction emerged after the PhD research. In many instances there would have been no direct interactions without financial interactions (i.e. financial support to attend specialist meetings). More often, the combination of direct and indirect interactions enabled a process of general stakeholder enlightenment characterised by "knowledge creep" (Weiss, 1978, p. 23). In the words of HF:

> "*You have ten people talking. They* [users] *may take one aspect from each of the ten people because, you know, you are influenced by what you hear, by the events that unfold. So you can't say, ah, because of that there this is, no, it's collective and progressive as well, as it goes along*."

Lastly, self-documentation as a form of data collection in research impact assessment, although valuable, needs to be supported by interview data.

References

De Jong, S., Barker, K., Cox, D., Sveinsdottir, T., & Van den Besselaar, P. (2014). Understanding societal impact through productive interactions: ICT research as a case. *Research Evaluation*, *23*, 89–102.

Fouché, H.F. (2006). *Policing piracy and armed robbery of ships in South Africa's territorial waters and contiguous zone*. Unpublished D.Tech dissertation, Tshwane University of Technology, Pretoria, South Africa.

Molas-Gallart, J., & Tang, P. (2011). Tracing 'productive interactions' to identify social impacts: An example from the social sciences. *Research Evaluation*, *20*(3), 219–226.

SIAMPI. (2013). *Final report on social impacts of research* (SIAMPI = Social impact assessment methods for research and funding instruments through the study of productive interactions between science and society). Retrieved May 14, 2017 from: http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf

Spaapen, J., & Van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation*, *20*(3), 211–218.

Weiss, C.H. 1978. Broadening the concept of research utilization. *Sociological Symposium*, *21*, 20–33.

Drowning by Numbers: Evaluating Social Capacities

Eiríkur Smári Sigurðarson

Impact and evaluation of research in the humanities has been the subject of several recent publications (e.g. Benneworth et al. 2016, Ochsner et al. 2016). A common theme is the need for a reevaluation of how the humanities are evaluated, as well as for research on the effects of evaluation practices on the humanities (Ochsner et al. 2016, p. 9). Another important idea is that this reevaluation should take into account the nature of the public value of the humanities (Benneworth et al. 2016, especially ch. 7). The dominant model of research evaluation has as its focus societal economic benefits of research, a focus humanities scholars have mostly been uncomfortable with (Benneworth et al., ch. 2). One of the main tools of research evaluation, bibliometric analysis, is problematic – to say the least – for the humanities (Ochsner et al. 2016). As the policy and practice of research evaluations influences the way research is conducted (van Leeuwen 2016) there is a potential danger that the inherent value of the humanities will be eroded and their value to society diminished *through evaluation*. This problem will be addressed from two sides in this paper.

The first approach is to look closer at the definition of the public value of research in the arts and humanities proposed in Benneworth et al. (2016, p. 199):

the circulation of research in networks to users with identifiable interactions creating things that make a good society as public benefits from private assets

A fundamental idea behind this definition is that of "social capacities" developed through humanities research. There is an important *local* aspect to this definition, recognized e.g. by UNESCO in its program on Social Transformations (UNESCO 2017). According to UNESCO the value of humanities (through research and teaching) for positive social transformations is found in the capacities developed in individuals and societies to deal with external and internal changes (e.g. climate change, poverty). The public benefit in this case is not economic but a better functioning society overall. One effect of current evaluation practices, often recognized, is that an increasing number of publications is in English. This is potentially harmful for societies – especially in developing nations – and goes against one of the most important values of the humanities: its local importance for developing social capacities.

The second approach is to look at this issue through the lens of virtue epistemology and especially through the work of Linda Zagzebski (Zagzebski 1996 and 2012). The Icelandic philosopher and former rector of the University of Iceland has argued that the main responsibility of a university is to develop epistemic virtues in academics and students and thereby also in societies (Skúlason 2015). In this part I will explore its relation to the idea of social capacities with the aim of clarifying the latter concept. The value of this approach to clarifying social capacities is in its focus on *epistemic* virtues, pointing more clearly to research and education.

The final aim is to clarify the notions of social capacities and epistemic virtues and thereby to further the debate on how to properly evaluate the humanities.

Acknowledgements

This research is developed within the ENRESSH Cost Action (enressh.eu).

References

Benneworth, P., Gulbrandsen, M., & Hazelkorn, E. (Eds.) (2016) *The Impact and Future of Arts and Humanities Research*. London: Palgrave Macmillan.

Ochsner, M., Hug, S.E. & Daniel, H-D. (Eds.) (2016) *Research Assessment in the Humanities*. Springer: Zurich.

Skúlason, P. (2015) "The University and the Ethics of Knowledge", in *Critique of Universities*. Háskólaútgáfan: Reykjavík.

van Leeuwen, T. (2016) "Philosophical and theoretical considerations on bibliometric analysis." Presentation at the 21st Nordic Workshop on Bibliometrics 3 November 2016. Retrieved February 27, 2017 from: https://figshare.com/articles/Philosophical_and_theoretical_considerations_on_bibliometric_analys is_NWB_2016_presentation_slides_/4248830

UNESCO "Social Transformations". Retrieved February 27 from: http://en.unesco.org/themes/socialtransformations

Zagzebski, L.T. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.

Zagzebski, L.T. (2012) *Epistemic Authority. A theory of Trust, Authority, and Autonomy of Belief*. Oxford: Oxford University Press.

Mapping the societal impact of SSH – a literature review

David Budtz Pedersen, Jonas Følsgaard Grønvad and Rolf Hvidtfeldt

The scientific and political agenda for the societal impact of social science and humanities (SSH) is not easily summed up. SSH is a heterogeneous field that consists of researchers with different motives, practices, and research styles working with different sectors in society. Currently, there is no agreement on what is the best definition or the superior methodology for assessing impact and knowledge exchange activities in SSH. Rather, several more or less overlapping frameworks for understanding, mapping, and evaluating SSH research impact compete for pole position. The impact agenda in SSH spans over several different literatures and institutions discussing various aspects of the societal relevance and uptake of research. As such, the literature on impact assessment and impact-stimulating initiatives is characterized by wide semantic diversification, including the often overlapping or synonymous use of concepts such as knowledge brokering, dissemination, co-creation, research partnerships, external engagements, research outcomes, uptake, mobilization and impact.

To create a better overview of the diverging literature, this paper aims to map both the scientific and political literature regarding the relevance, impact and uptake of SSH knowledge in society order to create a more substantial understanding of the current policy and academic agenda surrounding SSH. The paper addresses the following questions: How can we understand the societal relevance e.g. broader impact, social impact or public value of SSH research, and how might we meaningfully measure it? These questions are addressed through a systematic review of the current scientific literature combined with a survey of reports and policy papers from European research institutions. The paper will first provide a semantic analysis and map of keywords and concepts introduced in literature as a whole. Secondly the paper will review and compare established and selected approaches and models used to measure the societal relevance of SSH research. Finally, the paper points out issues which are inadequately dealt with in the existing literature and presents how we attempt to address these issues in an ongoing explorative impact assessment project at Aalborg University.

Method: Literature review and search strategy

The literature search for the survey was conducted at Aalborg University during 2016. The study selection process was divided into different sequential steps: A systematic search of publications (journal articles, monographs, reports etc.) was performed using computerized literature databases and internet search engines on selected keywords. The databases chosen for the review were SCOPUS and Web of Science. The database and web-search were followed up by tracing references in documents included in the review study. Additional documents were found through experts and peer consultations. In between these search steps,

screenings of selected documents were conducted applying the inclusion and exclusion criteria to all retrieved text titles, then abstracts and finally by reading and analyzing the

documents. Documents were only included in the final text corpus if they met all inclusion criteria. The selection process and empirical basis of the analysis is summarized in the flow-chart in figure 1.

Analysis 1: Semantic mapping of the SSH societal impact literature

In the paper we map the different, partly overlapping concepts that are used throughout the literature to characterize SSH research and that forms the conceptual basis for the discussions and assessment models put forward. To map the literature we use graph and network based methods to visualize the relationship between keywords in order to find influential and closely related concepts in the text corpus (Leydesdorff 2014; Carley, 1993). To identify the keywords we performed a qualitative deep reading and coding of the included texts. On this background we use text mining techniques to extract the selected keywords from all included texts and visualize the relationships between the key concepts.

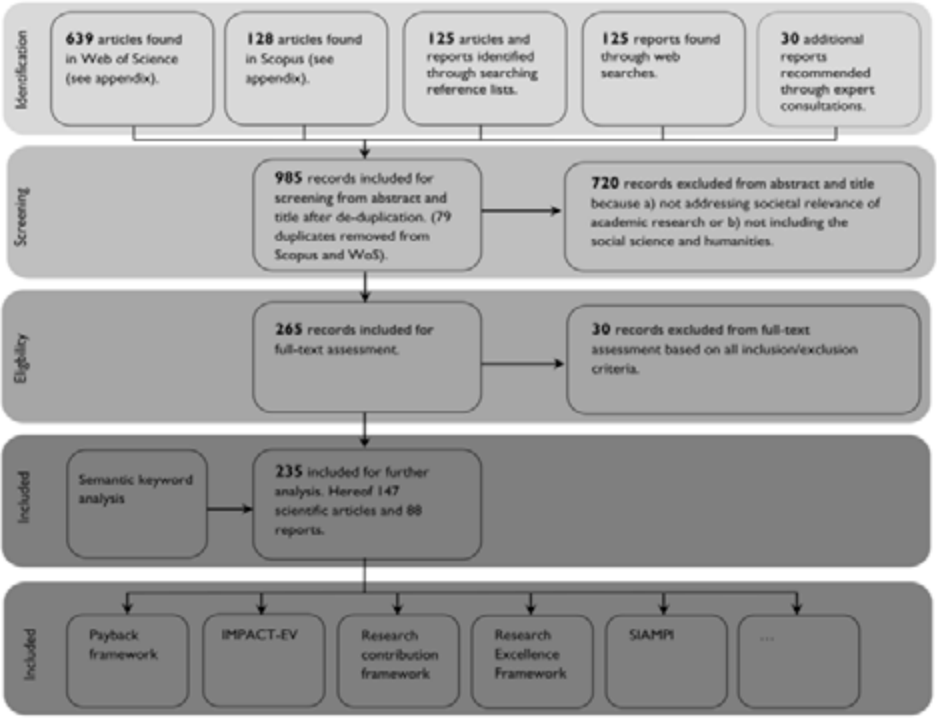
Analysis 2: Review of establish assessment models used on SSH research

The paper discusses the strengths and limitations of selected approaches used to assess SSH research. Here we look at similarities and differences between approaches and how they; define impact, understand the mechanisms by which impact is achieved, and the implications for how impact can be documented and measured. The models that will be analyzed include: The payback Framework (Levitt, Celia, & Diepeveen, 2010; Klautzer et al. 2011), the SIAMPI model (Spaapen & van Drooge, 2011; Molas-Gallart & Tang, 2011; Olmos-Peñuela, Molas-Gallart, & Castro-Martínez, 2014), Flows of knowledge framework (Meagher, Lyall, & Nutley, 2008), Research Contribution Framework (Morton, 2015), Contribution Mapping (Kok & Schuit, 2012), IMPACT-EV (Flecha et al., 2014) and The Research Excellence Framework.

Acknowledgements

The results presented in this paper are based on research conducted with support of the European Commission Horizon 2020 project ACCOMPLISSH, which aims to create a European platform for accelerating and mapping impact-oriented research in the social sciences and humanities.

Figure 2. Flowchart of the data selection process



References

Carley, K. (1993). Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis. *Social Methodology*, 23, 75-126

Flecha, R., Soler, M., Oliver, E., Puigvert, L., Sordé, T., Schubert, A., … Donovan, C. (2014). *Impact evaluation of FP6 (last call) and FP7 SSH research projects. Report 3*. *IMPACT-EV*. Retrieved from http://impact-ev.eu/wp-content/uploads/2015/08/D3.2-Report-3.-Impact-evaluation-of-FP6-last-call-and-FP7-SSH-research-projects.pdf

Klautzer, L., Hanney, S., Nason, E., Rubin, J., Grant, J., & Wooding, S. (2011). Assessing policy and practice impacts of social science research: the application of the Payback Framework to assess the Future of Work programme. *Research Evaluation*, *20*(3), 201–209. https://doi.org/10.3152/095820211X13118583635675

Kok, M. O. M. O., & Schuit, A. J. A. J. (2012). Contribution mapping: a method for mapping the contribution of research to enhance its impact. *Health Research Policy and Systems*, *10*(1), 21. https://doi.org/10.1186/1478-4505-10-21

Frameworks for understanding the societal relevance of the humanities

Gunnar Sivertsen

The aim of this paper is to compare and discuss different frameworks for the understanding of societal impact of research by testing them on an empirical material of reported impact cases from humanities.

Examples of relevant frameworks for the understanding of the societal impact of research in the humanities are the Payback framework (Levitt, Celia, & Diepeveen, 2010; Klautzer et al., 2011), the SIAMPI model (Spaapen & van Drooge, 2011; Molas-Gallart & Tang, 2011; Olmos-Peñuela, Molas-Gallart, & Castro-Martínez, 2014), the Flows of knowledge framework (Meagher, Lyall, & Nutley, 2008), the Research Contribution Framework (Morton, 2015), Contribution Mapping (Kok & Schuit, 2012), and the IMPACT-EV (Flecha et al., 2014).

These and other frameworks will be discussed in relation to more than 300 reported cases of societal impact that were included in the Research Excellence Framework (REF) in the United Kingdom in 2014 and in a similar national research assessment exercise in Norway in 2016. Comparison is possible because the Norwegian impact evaluation methodology was adopted from the REF.

In our view, the methodology for collecting and evaluating the reported cases of impact in the two countries is implicitly based on a certain framework for the understanding of societal impact that reminds of the so-called *linear model* of innovation (Godin, 2006) or communication (Shannon & Weaver, 1949). Amongst other things, the template for the case reports (REF2014, 2012) demands identification and documentation of:

- The research that *underpinned* the impact: "This section should outline the key research insights or findings that underpinned the impact, and provide details of what research was undertaken, when, and by whom,"
- The *resulting* impact: "A clear explanation of the process or means through which the research led to, underpinned or made a contribution to the impact (for example, how it was disseminated, how it came to influence users or beneficiaries, or how it came to be exploited, taken up or applied)."

The typical analysis of the case studies based on this model has been to identify *pathways*, *beneficiaries* and *effects* of research in the reported cases.

Mentioned above are examples of other frameworks for understanding impact which are more or less alternatives to the linear model. The production and use of knowledge is instead understood as a process of interaction and co-creation. Nevertheless, our preliminary tests of these other frameworks on the same reported cases indicate that they are also applicable and can provide alternative or supplementary understanding.

This paper will also investigate a possible expansion of the existing frameworks of understanding. We are interested in observing humanities research as integrated in, and not

operating at a distance from, certain domains in society where the disciplines have specific purposes and play specific roles.

By comparing cases from two different countries, we expect to see patterns of interaction with society that are typical for each of the disciplines and similar across countries. If this is confirmed, the alternative framework for understanding may also potentially influence the evaluation methodology. There could be a shift of focus from individual cases of demonstrated impact to a focus on the quality of the procedures with which the discipline continuously interacts with society at an organizational level.

References

Flecha, R., Soler, M., Oliver, E., Puigvert, L., Sordé, T., Schubert, A., … Donovan, C. (2014). Impact evaluation of FP6 (last call) and FP7 SSH research projects. Report 3. IMPACT-EV. Retrieved from http://impact-ev.eu/wp-content/uploads/2015/08/D3.2-Report-3.-Impact-evaluation-of-FP6-last-call-and-FP7-SSH-research-projects.pdf

Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. Science, Technology & Human Values, 31(6), 639-667.

Klautzer, L., Hanney, S., Nason, E., Rubin, J., Grant, J., & Wooding, S. (2011). Assessing policy and practice impacts of social science research: the application of the Payback Framework to assess the Future of Work programme. Research Evaluation, 20(3), 201–209.

Kok, M. O. M. O., & Schuit, A. J. A. J. (2012). Contribution mapping: a method for mapping the contribution of research to enhance its impact. Health Research Policy and Systems, 10(1), 21.

Levitt, R., Celia, C., & Diepeveen, S. (2010). Assessing the Impact of Arts and Humanities Research at the University of Cambridge. Technical Report. RAND Corporation. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED510286

Meagher, L., Lyall, C., & Nutley, S. (2008). Flows of knowledge, expertise and influence: a method for assessing policy and practice impacts from social science research. Research Evaluation, 17(3), 163–173.

Molas-Gallart, J., & Tang, P. (2011). Tracing "productive interactions" to identify social impacts: an example from the social sciences. Research Evaluation, 20(3), 219–226.

Morton, S. (2015). Progressing research impact assessment: A "contributions" approach. Research Evaluation, 24(4), 405–419.

Olmos-Peñuela, J., Molas-Gallart, J., & Castro-Martínez, E. (2014). Informal collaborations between social sciences and humanities researchers and non-academic partners. Science and Public Policy, 41(4), 493–506.

REF2014 (2012). Assessment framework and guidance on submissions. Retrieved from http://www.ref.ac.uk/pubs/2011-02/

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. Urbana, Illinois: University of Illinois Press.

Spaapen, J., & van Drooge, L. (2011). Introducing "productive interactions" in social impact assessment. Research Evaluation, 20(3), 211–218.

Towards a common understanding on the societal impact of SSH research

Paul Benneworth, Julia Olmos-Peñuela and Reetta Muhonen

Introduction

There has been an increasing set of demands placed upon science – including the social sciences and humanities (SSH) – to demonstrate the returns that science produces (Martin, 2011). What was once taken for granted fifty years ago – that investing in research was self-evidently worthwhile (Plumb, 1963) – has fallen prey to demands to be able to concretely demonstrate the value-added created (Nussbaum, 2012; Benneworth, 2015). There is a widespread recognition that the measures that have subsequently been developed for societal impact of research lag far behind the quality of those that have been developed via bibliometric to demonstrate scientific impact of research (Collini, 2009; Olmos-Peñuela et al., 2015). However, these 'societal impact' (hereafter impact) metrics to date –clustered around commercialisation outcomes – do enjoy a degree of legitimacy, and in the UK, for example, substantial funding is allocated to them via the Higher Education Innovation Fund (HEFCE, 2016).

It is much harder to make a similar claim for the relevance of these metrics to SSH disciplines, in part because of the much higher use of non-contractualised forms of knowledge exchange (Olmos-Peñuela et al., 2014a, 2014b). Thus, ease of measurement has become a proxy for the quality of the indicator, and from SSHs poor performance, and apparent reluctance to embrace these indicators, a series of negative inferences are drawn about the value of SSH to deliver socio-economic outcomes (Molas-Gallart, 2015). In the absence of other compelling evidence, SSH is therefore compelling to demonstrate its relevance by subordinating itself to other disciplines whose relevance is not in question. This may be for example being seen as a means of providing content for digital media technology companies, or in the level of research, in dealing with questions of societal acceptance for new technologies (Else, 2013; KancewiczHoffman, 2013; Benneworth et al., 2016).

Our diagnosis of this problem is that it arises because of a desire for simplistic definitions of valorisation and impact creation arising from policy-makers, and the resultant confusion that this has produced in SSH communities. The message has been that SSH produce many different kinds of impact and they are all in some ways valuable even if that value is not easily enumerated (Lee, 2013; Olmos-Peñuela et al., 2015). Any attempts to move further than that are met with resistance from academics who point to the irrelevance of these simplistic metrics (Collini, 2011). But in then translating that message to policy communities, we see that despite a language stating that SSH can be important, evaluation frameworks assume that the pathways that really matters are those impact that are most immediately relevant to the science, technology engineering and mathematics (STEM) disciplines (see for example the UK's pathways to impact approach). The question then arises of how to accommodate that diversity without creating a diffusion of definitions that ultimately lead to confusion, dissent and policy irrelevance.

To address this debate, we diagnose the problem as arising from a tendency within SSH to create hermeneutic understandings of impact, and to dwell on their idiosyncrasies, whilst

failing to adequately develop a wider deductive understanding of the common definitions of impact. At the same time, this diagnosis of a desire for hermeneutic impact narratives explains why there is such resistance to impact definitions, because they are compared to a hermeneutic standard – that is to say they cannot explain the complexity of the contours of individual cases. To move beyond a hermeneutic approach to understanding SSH impact, we therefore propose to develop a set of deductive propositions about SSH impact which highlight the most important common features. At the same time, such an approach must be candid about its ontological framing and acknowledging that these propositions (stylized facts) cannot possibly capture the diversity within the field, in particular in terms of the meanings and values that scholars ascribe to their own impact activities.

In this paper, we therefore ask the question about what are the key concepts, structures and relationships necessary to create a deductive structure of the field of SSH impact. We do so by attempting to understand the way that SSH in Europe define the creation of research impact, its relation to the impact process, its desirability, the tensions it creates, the underlying policy imperatives and its potential problems. From that, we seek to create a logical structure about the commonalities around SSH impact as the basis for further inquiry into impact practices which can lead to a convergence of ideas, without replicating the past tendency in SSH impact studies to divergently emphasise difference.

Methodology

To operationalise this, we note that the structure of SSH in Europe is often organised around national communities, and in particular, discussions about research impact are national debates (even where researchers are active in international research communities). We therefore choose to take ourselves the responsibility for rendering these perspectives into a common lexicon, and therefore choose for a questionnaire-based approach in a single language (English) of self-identified interested stakeholders in SSH impact. The experts were selected through their participation within a Working Group (WG) of the COST ENRESSH network (European Network on Research on Evaluation of the Social Sciences and Humanities).

The questionnaire was developed using a co-creation methodology with a selected group of experts within the COST WG network and then refined by circulating it to get comments from all participants. The final questionnaire included a total of 13 open questions each of which contained a number of prompts, as well as a final open question allowing further remarks, comments on the questionnaire itself as well as anything that the respondent felt necessary. A total of 29 questionnaires were returned, from 15 countries and from a broad range of fields within SSH, as well as from stakeholders who came from non-research positions within universities, science councils and other participants. These 29 returns form the basis for the analysis presented in the paper.

Preliminary results

The research findings highlight a number of commonalities, as well as a huge amount of divergence, diffusion and difference in the definitions and understandings in use in the

field. A common definition emerged around contributing to processes of social development through ongoing interactions between scholars and societal partners as being the most important mechanism despite the high profile of a number of 'heroic' academic scholars. Societal partners were necessarily defined very broadly, and in short reflected all those who interacted with the knowledge created by researchers who did not then use that knowledge to write additional scientific papers or advance research. There was agreement that even though SSH is diverse, it is homogenous in the sense of having the capacity for mutual respect and appreciation (for example between deductive and hermeneutic fields) in a way that STEM did not have for SSH.

There was a commonality of the models for creating impact in seeking to capture these ongoing interactions with society, with these interactions also having a shaping influence on the directions made in research choices (for example when specific societal problems emerged). Respondents (including policy-makers) agreed that policy makers' definitions of SSH impact at best gave SSH scholars the opportunities to try to stake a claim for their research's value and at worst were irrelevant. SSH researchers faced the additional challenge that engagement was not always positive for their research or for society and so had to actively manage and be aware of that process in the absence of societal consensus as to what good SSH impact constituted (analogous to spin-off companies). Also noted was that although everyone knew that not all academics should engage, there was strong societal negative pressure on those who did not. The paper concludes by arranging these propositions into a convergent definitional framework for SSH.

Acknowledgements

This article is based upon work from COST Action ENRESSH, supported by COST (European Cooperation in Science and Technology). We would like also to acknowledge the contribution of the working group participants who completed the questionnaire that fed into that study, as well as their feedback at other stages of the process as outlined above. Any errors or omissions remain the authors' responsibility.

References

Benneworth, P. (2015). Tracing how arts and humanities research translates, circulates and consolidates in society. How have scholars been reacting to diverse impact and public value agendas? Arts and Humanities in Higher Education, 14(1), 45-60.

Benneworth P. Hazelkorn, E., & Gulbrandsen, M. (2016). The impacts and future of arts and humanities research. London: Palgrave.

Collini, S. (2009). Impact on humanities: Researchers must take a stand now or be judged and rewarded as salesmen. Times Literary Supplement (5563), 18-19, 13th November 2009.

Collini, S. (2011). What are universities for? London: Penguin Books.

Else, H. (2013). Horizon 2020 Provision for humanities criticised. Times Higher Education, 5th December. Accessed 29th October 2014, from: http://www.timeshighereducation.co.uk/news/horizon-2020-provisions-for-humanities-criticised/2009603.article.

Garland, R. (2012). The humanities: plain and simple. Arts and Humanities in Higher Education, 11(3), 300-312.

HEFCE (2016). Key performance indicators. Accessed 2nd March 2017, from: http://www.hefce.ac.uk/kess/hebci/indicators/.

Kancewicz-Hoffman, N. (2013). The role of humanities and social sciences. Interview with Horizon 2020 projects. Accessed 25th February 2016, from: http://horizon2020projects.com/sc-society-culture-interviews/the-role-of-humanities-and-social-sciences/.

Lee, P. (2013). Declaration urges focus on social sciences, humanities. University World News, 5th October. Accessed 29th October 2014, from: http://www.universityworldnews.com/article.php?story=20131004141016572.

Martin, B. R. (2011). The Research Excellence Framework and the'impact agenda': are we creating a Frankenstein monster?. Research Evaluation, 20(3), 247-254.

Molas-Gallart, J. (2015). Research evaluation and the assessment of public value. Arts and Humanities in Higher Education, 14(1), 111-126.

Nussbaum M.C. (2012). Not for Profit: Why Democracy Needs the Humanities. New Jersey: Princeton University Press.

Olmos-Peñuela, J., Benneworth, P., & Castro-Martínez, E. (2014a). Are 'STEM from Mars and SSH from Venus'?: Challenging disciplinary stereotypes of research's social value. Science and Public Policy, 41(3), 384-400.

Olmos-Peñuela, J., Molas-Gallart, J., & Castro-Martínez, E. (2014b). Informal collaborations between social sciences and humanities researchers and non-academic partners. Science and Public Policy, 41(4), 493-506.

Olmos-Peñuela, J., Benneworth, P., & Castro-Martínez, E. (2015). Are sciences essential and humanities elective? Disentangling competing claims for humanities' research public value. Arts and Humanities in Higher Education, 14(1), 61-78.

Plumb, J.H. (1963). Introduction. In: Plumb, J.H. (Ed.) Crisis in the Humanities (pp.7-11). London: Pelican Originals.

Science Europe (2014). The Human Factor in the 2014–2015. Work Programme of the Horizon 2020 Societal Challenges, Humanities Scientific Committee Opinion Paper. Brussels: Science Europe. Accessed 29th October 2014, from: http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/SCsPublicDocs/Humanities%20Paper_FIN_dig.pdf.

Bibliometrics of SSH

Ostensible internationalization of journals in the social sciences and humanities in Poland as a result of the Polish Journal Rankings

Emanuel Kulczycki, Ewa A. Rozkosz, Aneta Drabek

Poland as one of the Central and Eastern European countries has undergone various transformations in academia in recent. After the breakdown of the Communist Regime, important systemic changes in university governance and funding modes were implemented. These changes have resulted in massive expansion of higher education institutions, while at the same time, Polish social sciences and humanities have lost significant amounts of international research visibility, and the so-called hard sciences have sustained their international research visibility.

This presentation discusses the transformations of Polish journals (N = 801) in the 2012–2015 period caused by the Polish Journal Ranking which is a key element of the Polish performance-research funding system. We have focused on the internationalization of journals in the social sciences and humanities (SSH), with the goal of investigating how science policy has transformed editorial practices at journals.

In our study, we have used one-way analysis of variance and two-way mixed design analysis of variance. Moreover, we have analyzed 15 semi-structured interviews. We selected four internationalization parameters from all parameters used in the journal evaluation process in Poland: the percentage of (1) authors from foreign countries, (2) reviewers from foreign countries, (3) articles published in the so-called congress languages, i.e. English, German, French, Spanish, Russian, and Italian, and (4) members of the editorial advisory board from foreign countries. Our analysis has revealed that science policy (more specific: research evaluation system) has transformed editorial practices, but that there is no actual internationalization in Polish social sciences and humanities journals. Rather, there is only the ostensible internationalization that manifests in "gaming" the journal evaluation system. We found that the editors of Polish journals do not discuss the challenges of internationalization, and implement only those internationalization practices that are explicitly required in the system regulations.

Our findings show that in general, science policy has transformed editorial practices in Polish journals through the establishment of internationalization parameters. Nonetheless, the internationalization that the parameters were intended to bring about has not become a reality, and there is no actual internationalization of journals in the SSH. There is only the ostensible internationalization that has been manifested in two changes within the editorial practices in Polish SSH journals. The first change is a considerable increase in the percentage of editorial advisory board members and reviewers from other countries and non-considerable increase of two others parameters. The other change is a minor increase in the percentage of authors from other countries. The ostensible character of internationalization is revealed by various practices: articles written in the so-called congress languages and published in Polish journals are distributed only to the local audience, the values of

internationalization parameters are adjusted (e.g. by adding or removing editorial advisory board members) solely for the purpose of obtaining more points in the research evaluation game. In the interviews, editors characterized such editorial practices as gaming.

The ostensible internationalization of Polish journals shows that the so-called Campbell's Law applies in this case. The internationalization parameters were intended to regulate and motivate the process of internationalization. However, they have become a target and a way of achieving goals other than those originally intended, i.e. obtaining the evaluation points.

The following conclusions relevant especially for policy makers can be drawn from our presentation. A research evaluation system might motivate the internationalization of journals, but it should deemphasize the corrupted parameters. As our findings show, journals in the SSH are strongly connected to the surrounding society and language. Moreover, a majority of such journals are local and devoted only to the local audience. Thus, the system of journal evaluation should acknowledge two types of journals in the SSH: those that are actually prestigious and internationally oriented or might become such, and those that are actually locally oriented and have implemented the highest standards of editorial practices. An updated and improved system of journal evaluation might encourage small journals (e.g. from a single university or region) to consolidate. Internationally-oriented journals should be motivated to increase their international audience, such as by publishing full texts in open access.

Representing the humanities in bibliometric scholarship

Thomas Franssen and Paul Wouters

Bibliometric indicators are increasingly influential devices through which the science system understands itself and based on which actors make decisions (Espeland & Sauder, 2007). Prior research on the constitutive effects of bibliometric indicators was primarily focused on individual researchers and organizations (De Rijcke et al., 2016). We are in particular interested in the ways in which the humanities are characterized in relation to other scientific domains and how bibliometric indicators play a role in establishing hierarchies of scientific domains.

We focus on bibliometric scholarship in the period 1965-2015 that studies the humanities empirically. We trace the development of measures, such as the Price Index, and the ways in which the humanities are conceptualized and operationalized in these measures. We understand the bibliometric data infrastructure as an 'experimental system' (Rheinberger, 2010 see also Wouters, 1999; 2006) which enabled the generation of new forms of knowledge about the sciences, social sciences and the humanities.

In the first part of the paper we focus on bibliometric scholarship between 1965 and 1979. In this period only four publications include the humanities in their empirical analysis. In these studies concepts such as hardness, cumulativeness and codification are used to develop an empirical analysis of differences between scientific domains. By analyzing the commonalities between these studies, we show how the humanities became embedded in a particular notion of what constitutes science and in a specific hierarchy of the scientific domains. Moreover, we argue that in these studies two different notions of the nature of the humanities are enacted. First, the humanities are conceptualized as consisting of group of subject fields and, second, conceptualized as a particular form of scholarship.

In the early bibliometric scholarship of Price and Storer, the first notion is used to operationalize the second. Both Price and Storer are looking for different 'moods' and 'metabolisms', and their point of departure is the humanities as a form of scholarship. They operationalize this by selecting journals from specific subject fields, assuming that all or most articles in these journals will be examples of this humanistic form of scholarship. This homogeneity between a mode of scholarship and particular subject fields is problematized by the end of the 1970s in the work of Cole, Cole and Dietrich. They find that subject fields are not homogeneous entities and therefore cannot represent a single form of scholarship.

The second part of the paper analyzes bibliometric scholarship from the 1980s onwards. In this period bibliometric measures become increasingly embedded in science policy and research evaluation practices. This causes a shift in the focus of bibliometric scholarship concerning the humanities. As data completeness becomes increasingly important, scholars move away from citation behavior and towards publication profiles (e.g. see Nederhof et al., 1988). The continued lack of coverage of humanities journals and books in major bibliometric databases has limited the development of citation theories in the humanities (but see Hellqvist, 2010). In the concluding section we reflect upon this history of bibliometric scholarships as well as future possibilities of bibliometric methods. We argue that

bibliometric methods hold important possibilities for humanities scholars, especially those in the emerging field of the history of humanities.

Literature

De Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. & Hammarfelt, B. (2016) 'Evaluation practices and effects of indicator use—a literature review.' *Research Evaluation*, 25(2), 161-169.

Espeland, W. N. & Sauder, M. (2007) 'Rankings and Reactivity: How Public Measures Recreate Social Worlds.' *American journal of sociology*, 113(1), 1-40.

Hellqvist, B. (2010) 'Referencing in the humanities and its implications for citation analysis.' *Journal of the American Society for Information Science and Technology*, 61(2), 310-318.

Nederhof, A.J., Zwaan, R.A., de Bruin, R.E. & Dekker, P.J. (1988) *Produktiviteit en kwaliteit in alfa- en gammawetenschappen; een haalbaarheidsonderzoek*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Rheinberger, H. J. (2010) *On historicizing epistemology: an essay*. Stanford University Press.

Wouters, P.F. (1999) *The citation culture*. PhD-thesis, University of Amsterdam, Amsterdam.

Wouters, P.F. (2006) 'What is the matter with e-Science? – thinking aloud about informatisation in knowledge creation.' *Pantaneto Forum*, no. 23, July 2006

Proposal of an assessment model for scholarly publishers applied to university scientific output

Elea Giménez-Toledo, Jorge Mañana-Rodríguez, Elba Mauleón-Azpilicueta, Daniela De Filippo and Elias Sanz-Casado

Introduction

The assessment processes of scientific output in place at Spanish Universities have been traditionally based on consolidated tools (WoS, ERIH, Scopus, etc.) concerning scientific journals. In the case books and book chapters, assessment methods have been heterogeneous in nature and, often, these have not been based on specific indicators or external expert panels, as it is common in several European countries (Giménez-Toledo *et al.*, 2016). In the institutional assessment processes, each university or research institution independently chooses the assessment criteria in order to prepare its annual reports, in order to provide salary incentives, or for the implementation of specific programs for the promotion of research.

Spain counts with the information system Scholarly Publishers Indicators, which provides over ten indicators for scholarly publishers, such as the perceived prestige by the research community, the specialization of the publisher in different fields (within its catalogue and also among other Spanish scholarly publishers), the manuscript selection procedures and the degree of production published in scientific or scholarly fields from the metadata provided by the publishers, and the presence in several international information systems.

Objective

The existence of information for scholarly publishers, which is already being used in individual assessment processes allows to proof and propose an assessment model for scholarly books (monographs, collective works and book chapters) based on quantitative and qualitative indicators, aligned with international practices and transparent for those under evaluation. That is the objective of this communication. The assessment model proposed has been applied to the output of Carlos III University (Madrid, UC3M hereafter) corresponding to the 2013-2014 period, as reflected in its research reports.

In the design of the model it is worth discussing the methodological decisions which can make the results significantly vary. It is expected to generate that discussion in the framework of RESSH 2017.

Methodology

The assessment method is applied to the output in the form of books and book chapters published by lecturers at UC3M in the 2013-2014 period, as included in the institutional CRIS which uses Universitas XXI's software. This communication would offer the analysis for the Journalism and Audiovisual Communication. From the available indicators for scholarly publishers, three have been selected since these are considered the most relevant concerning the evaluation of publishers: perceived prestige, percentage of output in scholarly fields and original manuscript selection processes.

Results

22 publishers have been identified in the analyzed output. The variables considered for the design of the model imply a set of methodological decisions which will be discussed:

a. Concerning publishers' perceived prestige: a publisher can occupy different positions depending on the ranking; it can be the general publishers' ranking or the ranking of the discipline in which the book has been published (in this case, Communication) or in the rankings of other disciplines and this implies that different positions with different weights can be considered (prioritizing, in example, the position in the ranking of the discipline.
b. Concerning the percentage of output in scholarly fields of the publisher, it is also possible to prioritize those publishers which output is mainly devoted to that gender; although generalist publishers might have collections specialized in scholarly fields, which would handicap their scores concerning this variable.
c. Concerning manuscript selection processes: the large diversity in these practices by scholarly publishers and the solid justification for each of them provided by the publishers, the transparency of publishers (towards authors) concerning the selection processes has considered a valuable feature, which implies higher score for publishers stating the assessment procedures they use. It is to say, relevance is given to transparency instead of prioritizing certain selection processes.

Book publishers have been classified in two categories, in line with the practices of other countries such as Norway, Finland, etc. Results tables would be included, in which the variability of the levels applied to the publishers (1 and 2) would be reflected according to the different possible methodological decisions taken.

Conclusions

The diversity of available indicators for scholarly publishers allows their classification in levels and the application of that classification to the output of an institution. In other countries expert judgment is nuclear to scholarly publishers' classifications. In the proposed model expert judgments are also involved, although not constituted as panels, since the perceived prestige is derived from an ample consultation to the Spanish scholarly community. This expert judgment is combined with other indicators accounting for publishers' activities. It is proposed a model in which the recognition of the publisher, its specialization and transparency has been prioritized, assigning weights in accordance to the value of each of the analyzed criteria.

This work proposes an assessment model as something that should be debated among specialists; that could improve the model with the inclusion of new consensual contributions, having as ultimate end the possibility of making comparisons of the quality of scientific output at the international level.

SSH journal publishing in Flanders and Finland

Janne Pölönen, Tim Engels, Raf Guns, Gunnar Sivertsen and Frederik Verleysen

Introduction

The objective of this study is to compare scholarly journal publishing in the SSH in Flanders and Finland with regards to the selection of journal and language for publishing, the coverage of the publications in Web of Science and Scopus, as well as collaboration patterns and development from 2011 to 2014. The study contributes to understanding of SSH publishing patterns across countries (van Leeuwen 2006; Ossenblok *et al*. 2012; Sivertsen 2016a), as well as possible effects of performance based research funding systems (PRFS) on SSH publishing (Hicks 2012; Ossenblok et al. 2012; Aagaard et al. 2015).

Data comes from national databases, VABB-SHW in Flanders and VIRTA in Finland, to which complete publication data from the universities' local current research information systems (CRIS) is integrated to support the PRFS. Journal articles in the VABB-SHW have earlier been compared to similar data in the CRISTIN database in Norway in 2005-2009 (Ossenblok et al. 2012). The VIRTA database has been studied for publishing patterns (Puuska 2014), but it has not been systematically compared with data from other countries.

In Flanders, the national funding-scheme from 2003 until 2009 included only journal articles published in Web of Science (WoS) indexed journals. This created an incentive for Flemish SSH researchers to increasingly use WoS journals as outlets in this period, and to actively pursue indexing of regional journals in WoS. In Norway, PRFS included also other peer-reviewed publications and provided no particular incentive in the SSH to select journals covered by WoS. The share of English language publications increased similarly in both countries (Ossenblok et al. 2012). Finland adapted the Norwegian model 2010-2012 (Gimenez-Toledo et al. 2016; Sivertsen 2016b).

Ossenblok, Engels & Sivertsen (2012) raised two questions that we want to address in this study. The first is the impact of the introduction of VABB-SHW in 2008-2010, which marked the inclusion of non-WoS-indexed peer reviewed publications in the Flemish PRFS. The second is the possible effect of whole counting in the funding-scheme on collaboration patterns in the SSH. In Norway, publications are fractionalized if more than one institution contributes to them. Both Flanders and Finland, however, use whole counting. As such, their systems could be said to incentivize inter-university collaboration because such articles count more than once in the funding system.

The analysis includes 11890 Finnish and 13145 Flemish articles that have been published 2011-2014 in SSH journals approved to be peer-reviewed in Flanders or Finland. Whole counts are used for both Flanders and Finland. SSH OECD FOS fields were identified for all journals in which Flemish and/or Finnish SSH researchers published in 2011-2014. Publication language data, where lacking or inaccurate, was manually checked. WoS and Scopus journals were identified on basis of ISSN for both datasets.

Results

The results are summarized in Figs. 1–10 and Table 1.

Figure 1. Share of publications in channels with publications from both Finland and Flanders, 2011-2014

Figure 2. Share of English language publications for Flanders and Finland, 2011-2014

Figure 3. Share of Web of Science publications for Flanders and Finland, 2011-2014

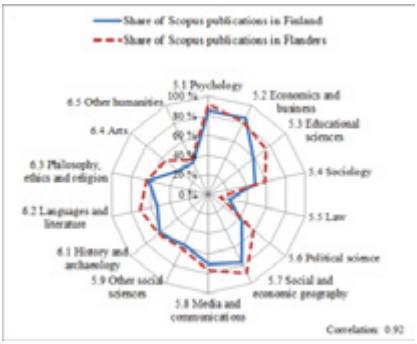


Figure 4. Share of Scopus publications for Flanders and Finland, 2011-2014

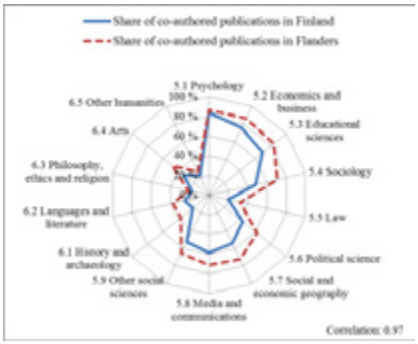


Figure 5. Share of co-authored publications for Flanders and Finland, 2011-2014

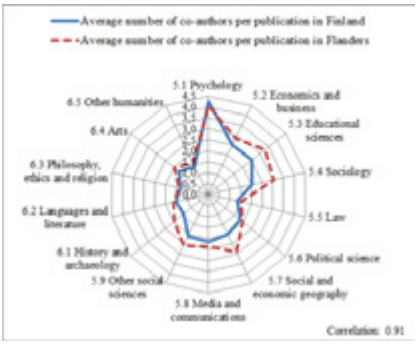


Figure 6. Average number of co-authors per publications for Flanders and Finland, 2011-2014

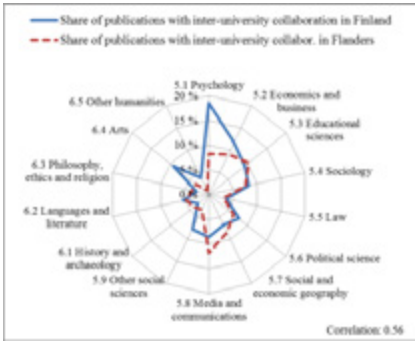


Figure 7. Share of inter-university collaborations for Flanders and Finland, 2011-2014

Figure 8. Share of English language publications in other than WoS-journals for Flanders and Finland, 2011-2014

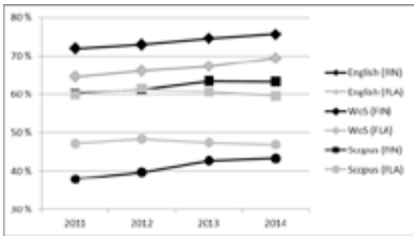


Figure 9. Share of English language, WoS and Scopus publications for Flanders and Finland, 2011-2014

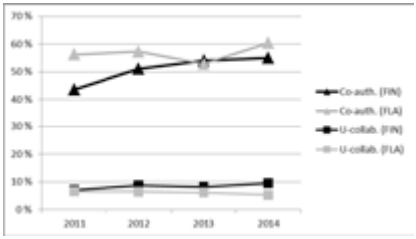


Figure 10. Share of co-authored and inter-university collaboration publications for Flanders and Finland, 2011-2014

Discussion

It has been shown on the basis of WoS data, as well as national publication data from Flanders and Norway, that different SSH fields show relatively similar publishing patterns across countries (Van Leeuwen 2006, Ossenblok et al 2012). This comparison of national journal article data from Flanders and Finland agrees with and contributes to these findings. Although there are overall differences in the shares of English language and co-authored publications between the two countries (Figures 2 and 5), the differences observed between SSH fields in Flanders and Finland are comparatively similar. Important deficits in WoS and Scopus coverage are also attested in quite the same fields (Figures 3, 4 and 8), because of which the international database coverage is not a proper measure of research quality or internationalisation in SSH (Sivertsen 2016a).

It can be expected that SSH researchers from Flanders and Finland do not publish in the same channels in instances where national or regional channels are more relevant for presenting research concerning national language, culture, history or society. Moreover, international publishing in the SSH is widespread in a variety of relatively small journals (Sivertsen & Larsen 2012). The comparison of journal publishing 2011-2014 indeed shows that in many SSH fields, the majority of Flemish and Finnish research is published in different channels (Figure 1).

English language publishing shows similar growth in both Flanders and Finland in 2011-2014 but in Flanders the share of WoS (and Scopus) publications has declined (Figure 9). It is tempting to see this development at least partly as the effect of including in the funding-scheme non-WoS publications in the VABB-SHW. In Finland, the share of WoS publications was considerably smaller to begin with but it has increased even if the funding-scheme (like that in Norway) is not tied only to WoS publications.

It seems that the funding-scheme may have little influence on the collaboration patterns. In Norway, fractionalization has not hampered co-authorship (Aagaard et al 2015). The share of both co-authored publications and inter-university collaborations is increasing in Finland (Figure 10). In Flanders, the share of co-authored publications was higher already in 2011 and has increased less rapidly, and the share of inter-university collaborations has slightly declined. There may, however, be interesting cultural differences between the conduct of SSH research in Flanders and Finland, as co-authorship is more common in all fields in Flanders and the average number of co-authors is larger (Figures 5 and 6).

References

Aagaard, K., Bloch C. and Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator, Research Evaluation, 24: 106–117.

Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T., Ingwersen, P., Pölönen, J., Sivertsen, G., Verleysen, F. & Zuccala, A. (2016). Taking scholarly books into account: current developments in five European countries, Scientometrics 107: 685–699.

Hicks, D. (2012). Performance-based university research funding systems, Research Policy, 41: 251–261.

Ossenblok, T., Engels, T. & Sivertsen, G. (2012). The representation of the social sciences and hu-manities in the Web of Science: a comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). Research Evaluation, 21: 280-290.

Puuska, H.-M. (2014). Scholarly Publishing Patterns in Finland: a comparison of disciplinary groups, Tampere University Press, Tampere 2014.

Sivertsen, G, & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential, Scientometrics 91: 567-575.

Sivertsen, G. (2016a). Patterns of internationalization and criteria for research assessment in the so-cial sciences and humanities, Scientometrics, 107: 357–368.

Sivertsen, G. (2016b). Publication-Based Funding: The Norwegian Model. In M. Ochsner et al. (eds.), Research Assessment in the Humanities: Towards Criteria and Procedures, Springer Interna-tional Publishing.

Van Leeuwen, T. N. (2006). The application of bibliometric analyses in the evaluation of social sci-ence research. Who benefits from it, and why it is still feasible, Scientometrics, 66: 133–154.

Alternative metrics

Towards values-based evaluation in the humanities and social sciences

Stacy Konkiel, Nicky Agate, Christopher Long, Simone Sacchi, Jason Rhody and Rebecca Kennison

For too long, humanists and social scientists have been allergic to metrics. This allergy has prevented researchers from engaging in a serious and sustained conversation about what practices of scholarship should be cultivated and incentivized both through the activities that are measured and those that are celebrated. As a result, a large and growing battery of metrics have been developed based on the practices of science scholars or simply on what it was possible to use our technologies to measure (Brooks, 2005; Hemlin, 1993; Hemlin and Gustafsson, 1996; Guetzkow et al, 2004; Oancea and Furlong, 2007).

Much focus has been paid to measuring usage and citations in the research literature. Applications of these metrics in humanities and social sciences (HSS) contexts, especially against the backdrop of efforts like the Research Excellence Framework in the United Kingdom, has generated considerable skepticism from scholars (Ochsner, Hug & Galleron, 2016; Wilsdon et al, 2015). Many HSS researchers believe that evaluation based only upon these metrics neither correctly assesses the impact of HSS scholarship nor serves the best interest of HSS researchers. Evaluators' collective challenge and responsibility, then, are to articulate, incentivize, and reward practices that enrich scholarly lives and expand the understanding of scholarship itself.

This paper, based on the work of the HuMetricsHSS initiative, proposes a fundamental change in how approaches to research evaluation in the humanities and social sciences are built. The paper calls for a holistic, values-based evaluation paradigm, one that uses metrics only to measure a scholar's progress towards embodying five values that initial research suggests are central to all humanistic and social science disciplines:

- Collegiality, which can be described as the professional practices of kindness, generosity, and empathy towards other scholars and oneself;
- Quality, a value that that demonstrates one's originality, willingness to push boundaries, methodological soundness, and the advancement of knowledge both within one's own discipline and amongst other disciplines and the general public, as well;
- Equity, or the willingness to undertake study with social justice, equitable access to research, and the public good in mind;
- Openness, which includes a researcher's transparency, candor, and accountability, in addition to the practice of making one's research open access at all stages; and
- Community, the value of being engaged in one's community of practice and with the public at large, and also leadership.

Most existing research that explores the links between evaluation, metrics, and values in the social sciences and the humanities addresses only the values of quality and originality (Ochsner, Hug & Daniel, 2016; KNAW, 2011; Hug & Ochsner, 2014; Hemlin, 1993; Gogolin, Astrom & Hansen, 2004; Guetzkow, Lamont & Mallard, 2004). Moreover, though

a survey of the literature suggests that academics and policymakers are well-aware of the gap between desired behaviors and related metrics, incentives, and evaluation practices, few institutions (especially in the United States and United Kingdom) are putting changes into place that can close that gap.

Where the HuMetricsHSS project differs from previous HSS evaluation critiques is in the exploration of a fuller suite of scholarly values, including and beyond the value of research quality. HuMetricsHSS also seeks to engage researchers and university administrators to create strategies to drive the adoption of recommended, values-based evaluation practices at the university-level.

This paper articulates scholarly values in depth and describes work currently underway to validate these values with HSS researchers themselves. Both traditional metrics (i.e., bibliometrics) and altmetrics are explored as means of measuring one's progress towards these values. Also described is the larger arc of the HumetricsHSS project being conducted by a team of scholars and information professionals working to find ways to expose, highlight, and recognize the important scholarship that goes into not only research activities, but also the all-too-hidden work of peer review, teaching, service, and mentoring.

References

Brooks, R. L. (2005). Measuring university quality. *The Review of Higher Education*, 29(1), 1-21. doi: 10.1353/rhe.2005.0061

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the social sciences and the humanities? *American Sociological Review*, 69(2), 190–212. doi:10.1177/000312240406900203.

Hemlin, S. (1993) Scientific Quality in the Eyes of the Scientist: a Questionnaire Study. *Scientometrics*, 27/1: 3–18.

Hemlin, S., & Gustafsson, M. (1996). Research production in the arts and humanities. A questionnaire study of factors influencing research performance. *Scientometrics*, 37(3), 417–432.

Gogolin I, Astrom F and Hansen A (eds) (2014) Assessing Quality in European Educational Research. Springer VS: Wiesbaden, Germany.

Hug SE and Ochsner M (2014) A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal of Education Law and Policy*; 10 (1): 55–68.

KNAW (Royal Netherlands Academy of Arts and Sciences). (2011). Quality indicators for research in the humanities. Amsterdam: Royal Netherlands Academy of Arts and Sciences. Retrieved from https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/quality-indicators-for-research-in-the-humanities.

Oancea, A., & Furlong, J. (2007). Expressions of excellence and the assessment of applied an practice-based research. *Research Papers in Education*, 22(2), 119–137.

Ochsner M, Hug SE and Daniel H-D (2012) Indicators for research quality in the humanities: Opportunities and limitations. Bibliometrie—Praxis und Forschung; 1, 4.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., … Johnson, B. (2015). *Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. *Higher Education Funding Council for England*. London.

Reader and Citation Counts of Journal Articles in Linguistics

Judit Bar-Ilan

Introduction

In this study, we intend to explore whether altmetrics, specifically reader counts, can contribute to the assessment of journal articles in linguistics. Linguistics was chosen, because it is considered among the most "science like" disciplines in the Humanities (Hicks, 2004).

Mendeley is a widely used free online reference manager, where users bookmark and download items to their personal Mendeley libraries. It is assumed that they intend to read or use or have general interest in these items. The number of Mendeley users who bookmarked and downloaded a specific item is aggregated and appears as the number of readers of the item, even though we are not sure that the users actually read the item, and what is their opinion about what they read. Since it is assumed that when an author cites a work in her paper, she reads the work it is interesting to find out whether there is a relationship between citing and reading. Several studies showed significant correlation of around .5 between reader and citation counts (e.g. Zahedi, Costas & Wouters, 2014; Haustein at al., 2014). More specifically a large-scale study of SSH papers indexed in WOS (Clarivate's Web of Science) conducted by Mohammadi and Thelwall (2014), showed that among the selected Humanities disciplines the correlation between citation and reader counts were the highest (r=.454), where the median number of citation was 2, and the median number of readers was 4, for articles published in 2008 and citation and reader counts retrieved in 2012.

An additional aspect that raises interest in Mendeley is its extremely wide coverage of academic articles, even though it is created by individuals saving items to their personal libraries. Mendeley has the highest coverage among currently used altmetric indicators (e.g., tweets, blog mentions, news media mentions). For example, Mendeley readers bookmarked 97% of the articles published between 2001 and 2011 in the leading journal in information science, JASIST (Bar-Ilan, 2012); and Mohammadi and Thelwall (2012) were able to locate on Mendeley 47% of the articles published in liguisitcs in 2008 and indexed by WOS. Altmetrics are early signals of impact, as described by the "altmetrics manifesto" (Priem, Taraborelli, Groth & Neylon, 2010) as they accumulate much faster than citations.

Data collection and results

The specific aim of this study was to compare reader counts with citation counts of peer reviewed journal articles and reviews in linguistics. Data on journal articles and reviews published in 2015 and 2016 were collected from WOS for the 25 highest ranked journals based on the impact factor (JIF) published in the JCR (Journal Citation Reports) for 2015. For each article both the citation counts and the number of Mendeley readers were recorded. Data from WOS was collected directly from the website, while Mendeley reader counts were collected using Mike Thelwall's Webometric Analyst (http://lexiurl.wlv.ac.uk/) and

were cleansed to correctly match reader counts with citation counts.

Altogether 2087 articles were located on WOS, out of which 1087 received no citations (52%) and 1000 articles/reviews that were cited at least once. On Mendeley there are either no readers of an item, or there is at least one reader – this is why we consider only cited items on WOS. On Mendeley only 316 (15%) articles/reviews could not be found, thus its coverage is 85%. There are 925 articles/reviews that are both cited and read (44% of total).

In the subset of the 925 articles/reviews that were both read and cited, the average number of citations was 2.6, and the average number of reads was 24.1 – nearly ten times more! For the 846 articles/reviews that were not cited but only read the average number of readers was 14.1, while for those only cited but not read (75 items), the average number of citations was 1.7. The number of items neither cited nor read was 241. These findings indicate the advantage of including Mendeley reader counts as supporting evidence in assessment of linguistics researchers and research units. The Spearman correlation between reader counts and citation was only .383 ($p<.01$), most probably because of the low citation rates. Further studies are in progress to cover a longer publication span, and to include other citation and altmetric sources. Table 1 below provides further evidence per journal to support the inclusion of Mendeley reader counts into evaluations in the field of linguistics.

Acknowledgements

Thanks to Mike Thelwall for developing and maintaining Webometric Analyst.

References

Bar-Ilan, J. (2012). JASIST@mendeley. In altmetrics12 workshop at the *ACM Web Science Conference 2012* . Retrieved February 11, 2017 from: http://altmetrics.org/altmetrics12/bar-ilan/

Hicks, D. (2004). The four literatures of social science. In *Handbook of quantitative science and technology research* (pp. 473-496). Springer Netherlands.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, *101*(2), 1145-1163.

Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, *65*(8), 1627-1638.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto*. Retrieved February 11, 2017 from: http://altmetrics.org/manifesto/

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, *101*(2), 1491-1513.

Table 1. Citations vs readership counts per journal

	Journal	# items	# cited	ave. no. cits	# read	ave. no. read	JIF (2015)
1	Journal of memory and language	117	82	2.67	116	25.26	5.218
2	Applied linguistics	65	39	3.51	59	27.85	3.250
3	Brain and language	168	105	2.94	167	26.37	3.038
4	Bilingualism-language and cognition	123	65	3.29	120	22.58	2.330
5	Studies in second language acquisition	47	28	1.79	43	17.37	2.234
6	Journal of fluency disorders	44	18	2.50	41	12.98	2.022
7	Computational linguistics	40	7	2.29	34	14.00	2.017
8	Language teaching	47	20	2.80	41	18.49	2.000
9	Language learning	77	43	3.02	70	25.34	1.869
10	Mind & language	52	20	1.90	16	12.63	1.867
11	Applied psycholinguistics	116	69	2.68	108	15.11	1.833
12	International journal of language & communication disorders	124	46	2.07	121	15.58	1.798
13	Language	50	19	2.00	21	10.00	1.750
14	Journal of second language writing	37	25	2.08	36	40.81	1.744
15	Language acquisition	32	10	2.70	23	8.52	1.735
16	Computer assisted language learning	93	39	2.18	91	16.25	1.722
17	Research on language and social interaction	46	25	2.28	34	9.00	1.628
18	Second language research	46	15	2.33	46	18.02	1.568
19	Journal of english for academic purposes	81	29	1.76	80	24.66	1.558
20	Journal of speech language and hearing research	278	110	2.10	205	12.20	1.526
21	Language in society	47	17	1.88	35	8.34	1.525
22	Tesol quarterly	82	40	2.08	56	22.14	1.513
23	Linguistic inquiry	43	9	2.22	25	4.96	1.511
24	Language cognition and neuroscience	162	83	3.08	117	20.47	1.470
25	Language teaching research	70	37	1.97	66	23.14	1.444

Usage-based indicators for the Social Sciences and Humanities in the Web of Science

Evangelia Lipitakis

In this presentation we discuss certain "interest" indicators that measure usage for all items in the Social Science Citation Index and Arts & Humanities Citation Index of the Web of Science. These usage indicators aim to be used in the process of research discovery and in this presentation we explore the concept of interest for Social Sciences and Arts & Humanities research worldwide and its implications when used to complement citation analysis.

Background

Citation analysis and bibliometric indicators have long been efficiently used for the quantitative evaluation of research work and peer-reviewed journals. Citations from a published item to another are considered to be a measure of impact, significance and utility of scientific work and have been used to support researchers and research organizations in the process of identifying influential literature, forming publication strategies, journal collection development and research performance evaluation.

One of the challenges of using citation based indicators for the assessment of scientific and scholarly impact is the differences that scientific and scholarly disciplines having in their citation behaviour patterns, mainly depending on the discipline's average reading turnaround, publication practices and citation rates. It has been proposed that in natural and life sciences the average peak in the number of citations is the third or fourth year cycle, while in the Social Sciences the time window is much longer, i.e. around fifth or sixth year cycle. This is evident when using citation counts in certain disciplines within the Social Sciences and Humanities, such as Web of Science categories Romance Languages, Rhetoric or Architectural History, were the citation rates are traditionally low or nominal and citation based indicators can be insufficient measures of scholarly impact.

Another limitation of citation based indicators is the inherent citation lag to a published item. The citation life-cycle theoretically begins the moment an article is published, however, for an article to be cited it has to be accessed, read, cited in another article's reference list, sent for publication, revised and then published. This can be problematic when researchers would like to obtain evidence of impact in recently published items, i.e. days, weeks or months after publication.

Recently, the academic community has shown an increased interest in newly developed article level methodologies based on usage metrics which purport to have the ability to capture patterns of immediate impact of a published item much earlier in the conventional citation life-cycle. As the scientific and scholarly literature discovery process is becoming widely adopted via online citation databases, the literature accessibility and usage statistics can be measured and calculated in real-time, while the users are actually navigating the online citation database. Usage counts are based on the concept of a researcher's intention

to capture a published item for acquiring information from this record for use at a later stage i.e. reading it and/or citing it, while navigating the online citation database. Item Level Usage Metrics (ILUMs) are becoming increasingly popular quantitative indicators that can be used to complement citation analysis for acquiring rapid feedback on the evaluation of recently published items, as well as helping the scientific community to formulate and answer basic research questions such as what researchers are reading today, what researchers are intending to cite in the near future and how scientific behavior is evolving on a day to day basis. Even though attention from the scholarly research community has been brought to the use of the new "intention" indicators, a group of advanced robust indicators have not yet derived from usage data other than publication download scores. Some challenging topics in measuring online usage counts is the complexity of the events and behaviours within the large scale network of aggregated metadata, privacy issues , the possibility of artificial inflation of statistics as a result of the automated extraction of information in the platform and distinction between human and robotic download traffic.

Methodology

Recently, a new class of Usage Count indicators that investigate the concept of interest towards a scholarly artifact are available for all items indexed in the Web of Science. The WoS Usage Count metrics can provide valuable quantitative analysis prior to, or in the absence of, citation activity and are based on a systematic quantitative analysis of the WoS users' literature retrieval behavior, by measuring actions such as access to the full-text or exporting content to bibliographic management tools. The WoS Usage Count indicators are calculated based on the worldwide daily logs of millions of WoS users' retrieval requests. Currently, there are two different Usage Counts indicators available in WoS, calculated based on two different time windows; (i) from present time to 2013 and (ii) for the last 180 days (moving time period).

Discussion

In this presentation we study the usage patterns in the Social Sciences Citation Index and Arts & Humanities Citation Index of the Web of Science. Fields and sub-fields with disciplinary variations and sparse citation rates are investigated to explore whether usage count metrics can be used to measure scholarly real-time, non-citation based impact in conjunction with citation counts. Analysis will be presented on topics such as citations and usage counts over time period for selected Social Sciences and Humanities fields, whether high interest papers become Highly Cited or Hot papers; Social Sciences and Humanities journals by average usage and average citations, usage patterns of open access versus traditional subscription papers, etc. Implications and limitations of the analysis will be discussed. As we launch these indicators and develop them and others further, we welcome discussion with the Social Sciences research community.

Toward alternative metrics of SSH research impact: A comparison of visit, download and citation data

Jadranka Stojanovski[1], Franjo Pehar[2], Tomislav Jagušt[3] and Draženko Celjak[4]

[1] jadranka.stojanovski@irb.hr
University of Zadar, Department for Information Science, F. Tuđmana 24i, 23000 Zadar, Croatia / Ruđer Bošković Institute, Library, Bijenička c. 54, 10000 Zagreb, Croatia

[2] fpehar@unizd.hr
University of Zadar, Department for Information Science, F. Tuđmana 24i, 23000 Zadar, Croatia

[3] tomislav.jagust@fer.hr
University of Zagreb, Department of Information Science, Unska bb, 10000 Zagreb, Croatia

[4] drazenko.celjak@srce.hr
University of Zagreb, University Computing Centre, J. Marohnića bb, 10000 Zagreb, Croatia

Introduction

Readers of the journal articles can express their interest, positive attitude, acknowledgment or tribute in different ways. The most traditional way is to cite the article in the own article's reference list. Nowadays' channels of the scholarly communications are changing, and journal articles are visited, downloaded, bookmarked, mentioned, commented, shared, and cited using different platforms and tools, social media and web applications (Donato, 2014). Indicators of the usage can be also visits and downloads, and in addition to the citations, can give us the more comprehensive picture about popularity and impact of journal articles. (Brody, Harnad, & Carr, 2006) The relationship between downloads and citations from a bibliometric perspective was investigated and possible differences in obsolescence patterns for articles published in a single journal were revealed. (Moed, 2005) The same study showed that, as a cohort of documents grows older, its download distribution becomes more and more skewed, and more statistically similar to its citation distribution.

Two criteria for academic careers are prevailing in the Croatian scientific community: number of papers published in the journals, and the prestige of the source journals that is measured, according to the discipline, with the Journal Impact Factor (JIF) (Macan & Petrak, 2015) and other qualitative indicators. In the social sciences and humanities the prestige of the journal is proved by the inclusion of the journal in the commercial databases like Scopus (Elsevier) and Web of Science Core Collection – WoSCC (Clarivate Analytics, former Thomson Reuters) etc., or the specific journal is recognized as high-quality journal by the community and listed as eligible for the assessment. Academic policy decision makers neglect the existence of other bibliometric indicators and the article level metrics. Because of the present criteria, authors are under pressure to publish their papers in high JIF journals. In the fields of social sciences and humanities researchers are often facing a dilemma to publish in prestigious, frequently fee-based, international journals or in local OA journals well recognized by the national scientific and professional community. The repository of the Croatian Open Access journals Hrčak includes 429 OA journals, and the majority of them come from the social sciences, humanities, and arts (n=302). In this fields research is more regionally and nationally engaged, and biased metric as JIF is particularly problematic (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015).

In this paper, the popularity and usage measured by a number of visits and downloads of the articles published in Croatian OA journals are compared with the number of citations from WoSCC and Scopus databases at journal and article level. The research question for this study is: Is there a correlation between metrics based on citations and metrics based on visits and downloads? According to the research question, the objective of our research is to investigate the relationship between visits and download data and traditional impact metrics such as citation analysis, in order to find out if the different methods are correlating or complementary.

Methodology

The usage data for 155.602 papers exported from 429 OA journal titles were provided by the repository of the Croatian OA journals Hrčak, and stored into a database with the following metadata: ID, article title in Croatian, article title in English, journal title in Croatian, journal title in English, ISSN, eISSN, volume, issue, publication year, date of publishing, date of the inclusion in Hrčak, number of visits, and number of downloads.

Even the Hrčak repository was launched in 2006, because of additional re-calculation of the usage data in 2013 which remove all robotic and non-human traffic, the usage data from Hrčak is available only from March 2007 to the present. Available Hrčak web usage statistics includes cover journal's visits, issue visits, article (bibliographic records) visits and full-text (PDF and HTML) downloads. For the purpose of this study article page visits and full-text downloads were considered. The article page visit with the bibliographic data includes a unique visit from the same IP address during the period of 30 minutes.

We identified 48 OA journals available in Hrčak repository and indexed by WoSCC and Scopus databases. Finally, we selected 17 journals from the field of social sciences and humanities. Citation counts for the articles published in these journals were obtained by using the print and online ISSN as the main search criteria. We exported 6.641 paper records from WoSCC and 4.482 papers from Scopus. Following fields from both databases were recorded in two separate CSV files: authors, source title, publication year, volume, issue, beginning page, ending page, and a number of total citations.

Further, we assemble a single data source based on three datasets. Since not all Croatian journals have DOI and there was not unique ID among three data sets, the data from Hrčak, Web of Science and Scopus were matched using the Levenshtein distance algorithm to determine the similarity between paper titles. In the first iteration all types of the papers were included: 'Book Review', 'Conference paper', 'Editorial', 'Essays', 'In memoriam, Obituary', 'Meeting abstract', 'News', 'Original scientific paper', 'Other', 'Preliminary communication', 'Professional paper', 'Reminiscences', 'Review article', 'Short communication, Note'. Further, only the papers with the citation potential were selected: 'Conference paper', 'Original scientific paper' and 'Review article'. This procedure in the last iteration matched 2.226 unique papers with data about visits and downloads from Hrčak and citation counts from WoSCC and Scopus.

The data was collected during the period 1 to 20 February 2017.

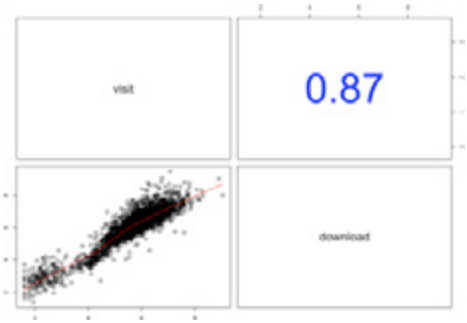


Figure 1. Hrčak total bibliographic record visits vs full-text downloads

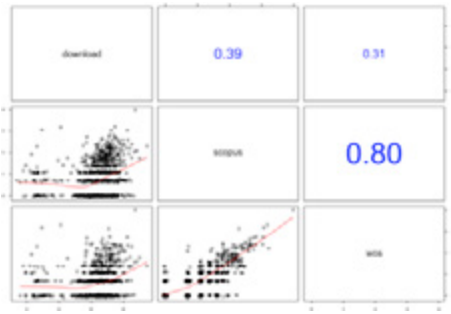


Figure 2. Hrčak full-text downloads vs Scopus and WoSCC citations

Results and discussion

The comparison between the total number of bibliographic record visits and full paper downloads is shown in Figure 1.[1] The correlation coefficient of the Hrčak bibliographic records visits and paper downloads shows high positive correlation (Spearman r=0.87, p=,00). The Spearman correlation results in Figure 2 show that there is a high positive correlation between Scopus and WoSCC citations (r=0.80) and medium to low correlation between full paper downloads and Scopus citations (r=0.39), as well as for WoSCC citations (r=0.31) for the same set of papers. These results show that full paper download frequency within the Hrčak repository weakly correspond to the Scopus and WoSCC citations.

1 We added 1 to all variables to enable it to be plotted on a log scale.

Conclusion

This paper presents the results of the study where different methods measuring usage (citations, visits, and downloads) has been applied. As described, there was medium to low correlation between the download data from Hrčak repository and citations from WoSCC and Scopus. This could imply that research of the regional or local relevance, published in the local journals and the local language, cannot gain a significant number of citations from the global community. Also, the usage patterns could be different between citing and downloading, related to the claim that downloads and citations measure different aspects of usage (Moed & Halevi, 2015). Citation databases as WoSCC and Scopus contain various biases such as under-representation of non-English journals (Jarvey, Usher, & McElroy, 2012), and the citations from non-indexed Croatian SSH journals are not included in WoSCC and Scopus.

The analysis of the relation between visits of the bibliographical records and the number of downloads demonstrated a high positive correlation, which could be the result of the links to the full-text article at the bibliographical level. The latter is important when discussing the advantages of open access publishing. There was also a high positive correlation between WoSCC and Scopus citations, which was expected according to the different studies (Archambault, Campbell, Gingras, & Larivière, 2009; Meho & Rogers, 2008). Since reading and citation populations are different (Moed & Halevi, 2015) the results could indicate the difference in reading and citing habits from the local community, and the international research community, where the visibility, readability, peer judgment, and citations depend on the English language, global relevance of the topic and prestige of the journal or author.

This study makes a contribution to the clarification of the question of the broad impact of altmetrics. Furthermore, these results raise questions regarding the appropriateness of the JIF and citations as a sole metrics in the assessment of article impact and suggest the possibility of devising impact metrics based on usage information in general. (Bollen, Van de Sompel, Smith, & Luce, 2005)

Limitations of the study

Hrčak serves as a national repository of Croatian OA journals, and about 70% of Hrčak journals have their own websites, with available bibliographic data and full-text articles. Because all articles are OA and licensing allow copies and redistribution of metadata and full-text, articles' PDFs are spread across various databases, repositories, and social networks. To get a complete picture about the number of downloads all these data should be aggregated.

References

Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the web of science and Scopus. *Journal of the American Society for Information Science and Technology*, *60*(7), 1320–1326. https://doi.org/10.1002/asi.21062

Bollen, J., Van de Sompel, H., Smith, J. A., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management*, *41*(6), 1419–1440. https://doi.org/10.1016/j.ipm.2005.03.024

Brody, T., Harnad, S., & Carr, L. (2006). Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Society for Information Science and Technology*, *57*(8), 1060–1072. https://doi.org/10.1002/asi.20373

Donato, H. (2014). Traditional and alternative metrics : The full story of impact. *Revista Portugesa de Pneumologia*, *20*(1), 1–2. https://doi.org/10.1016/j.rppneu.2013.11.001

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.

Jarvey, P., Usher, A., & McElroy, L. (2012). *Making Research Count : Analyzing Canadian Academic Publishing Cultures*. Toronto. Retrieved from http://www.higheredstrategy.com

Macan, B., & Petrak, J. (2015). Bibliometrijski pokazatelji za procjenu kvalitete znanstvenih časopisa. In I. Hebrang-Grgić (Ed.), *Hrvatski znanstveni časopisi : iskustva, gledišta, mogućnosti* (pp. 37–53). Školska knjiga, Zagreb.

Meho, L. I., & Rogers, Y. (2008). Citation Counting , Citation Ranking , and h -Index of Human-Computer Interaction Researchers : A Comparison between Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, *59*(11), 1711–1726.

Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, *56*(10), 1088–1097. https://doi.org/10.1002/asi.20200

Moed, H. F., & Halevi, G. (2015). On Full Text Download and Citation Distributions in Scientific-Scholarly Journals. *Journal of the Association for Information Science and Technology*, *67*(2), 412–431. https://doi.org/10.1002/asi.23405

Case studies per country

How Spanish book publishers select their monographs?

Elea Giménez-Toledo and Jorge Mañana Rodríguez

Introduction

For scientific journals, peer review is still the most accepted method for the validation of the scientific contents as well as an essential requisite in research evaluation. In the case of scholarly books, the requirement for the validation of the scientific content is implicit, but there are few studies on the original manuscript selection processes carried out by scholarly publishers. Sources indexing scholarly books accept those with peer review and exclude those without it, but there is a lack of information on the practices of each publisher.

Objectives

Considering that the manuscript selection practices of book publishers are not the same as those used in the case of journals, and assuming as well that the nature of the publishing projects do differ between multinational publishing companies, commercial publishers oriented towards the national market and / or university presses, this work aims at describing which are the manuscript selection processes of scholarly publishers in Spain and whether these differ among the various types of publishers. It is also the objective of this work to explore the reasons leading to the use of the different practices. Only by knowing how books are selected and which are the reasons for it, is it possible to extract conclusions on which aspects are to be considered, or are of specific value in the assessment processes of scientific output and, more specifically, monographs.

Methodology

The starting point is a study developed with the support of the Federation of Publishers' Guilds of Spain. It included a survey to Spanish scholarly publishers (commercial publishers, university presses and institutional publishers) concerning the processes carried out for the selection of manuscripts. The survey was sent between November, 2016 and January, 2017. The survey was sent to 505 publishers, with a 25% response rate (130 usable answers). The objective of the survey was wider but one of the questions addressed was the manuscript selection processes. More specifically, the respondents were asked to indicate which method is applied by the book publisher in order to select the manuscripts. The response options were: a) the opinion of the editor, b) the opinion of the book serie's director, c) the report of an internal reading committee (within the publishing house) and, d) the report of an internal reading committee (within the collection). The second related question referred to the transparency of the selection processes in the websites of the publishers and if, therefore, these are available to authors, potential authors and

Results

The analysis of the manuscript selection practices in the 130 publishers shows, mainly, diversity. Traditional peer review (experts in the field, external to the publisher) is the most common method (used by the 73% of the book publishers); nevertheless, it is not the only one. In general terms, it can be observed that a combination of methods and criteria are considered when taking a decision on the publication of a book.

Globally, the opinion of the book serie's director and the reports or opinions from internal reading committees are also widely used by the publishers. Around 55% of them use one of the two methods, in combination with others.

Differences are observed between the manuscript selection processes carried out by university presses and commercial/private publishers. In the case of university presses, peer review, together with the opinion of book serie's director and the internal reading committee is predominant; these three procedures are shared and / or have the same weight in the case of commercial publishers. However, in the case of commercial publishers, the opinion of the publishers' director has a prominent weight, which seems reasonable considering that in many cases commercial publishers are small or family business or are closely attached to the project of the editor, which risks its investments with each decision taken.

Conclusions and discussion

The results confirm the existence of a large diversity in the practices of manuscript selection among publishers. Not all decisions are taken according to the results of peer review and there are differences between the practices of university presses and commercial publishers. All this allows concluding that the publication of books is not similar –neither in this case- to the publication of journals. Apart from knowing the publishing dynamics related to the selection of manuscripts from among the large research output available, this study also shows that scientific validation of the results has to be accompanied by an assessment of the interest of the title for the market or the opportunity it represents when filling a publishing gap. That responsibility, which is closer to that of the publishing teams, is different from that of the journals. These issues should be known and considered when the activity and output of a scholarly publisher are assessed.

Impact of performance-based funding on publication patterns: Money changes everything...?

Jaroslav Šušol and Marta Dušková

Publishing (or publication) behavior is most prominently embodied in various ways and approaches that the authors apply when it comes to publishing, and especially to selecting and using diverse communication and publication channels. It is possible to explore the publishing behavior using various methods, quantitative and qualitative, analyzing the subjective interpretation of reality among authors themselves (questionnaires, surveys or interviews) or measuring hard data in the form of publication units bibliographic records.

One of the most important impulses for the development of publishing behavior research arrived in 1960s, with the application of computing technology into the processes of secondary information storage and retrieval. It gave birth to large bibliographic databases that were ready for quantitative analyses of various kinds. Research projects concentrated mostly on specific conditions and problems of particular scientific discipline.

Kyvik's research focused on assessing the evolution of publishing behavior within 2 decades and analyzed changes in publication patterns at the universities in Norway. He concluded that the extent of publishing oriented towards an international audience has increased and the scientific article in an international journal consolidated its position as a dominant type of publication (Kyvik 2003). At the turn of the century, the trustworthiness of electronic publishing became an important issue. In 2000, this kind of research took place at the universities in Canada (Rockwell 2000). The Swedish survey from the same year concentrated on the situation in building engineering and concluded that both the students and professors consider reviewing to be an important part of electronic publishing processes (Björk 2000).

One of the first major surveys of this type in Slovakia was done in 2002 by means of a questionnaire among the users of academic libraries. It focused, among other things, on the authors' relationship towards traditional and electronic publishing technologies. The results supported the assumption that the acceptance of electronic communication channels for publishing the output of scientific research depends on particular discipline, with strong unbalance between social sciences and humanities on one side and natural and technical disciplines on the other (Šušol 2004). A combination of qualitative and quantitative methods was applied in the survey of publishing preferences of PhD students at the Faculty of Arts, Comenius University in Bratislava. The results indicate, among other things, that the refusal of electronic publishing channels is caused rather by their non-acceptance in the authors' closest peer environment, not that much by the the authors' own attitude (Steinerová et al. 2006).

It is clear that the publishing behavior is a multi-dimensioned research domain, as the behavior of information process actors is a dynamic phenomenon influenced by various factors. The attitudes, opinions, preferences, that together define publishing behavior of authors in academic area, are not only subject to evolution in time, but have also certain local or national particularities. Of course, there are "natural", or historical differences

among various disciplines, as it was shown in many articles that tackled the topic from both quantitative (e.g. a very thorough statistical analysis and comparison of disciplines by clustering of authors through the analysis of their publication patterns by Verleyssen and Engels (Verleyssen et al. 2015)) or qualitative point of view (e.g. Steinerová in her series of interviews with researchers from various disciplines as they were carried out at the Comenius university in Bratislava during 2016 (Steinerová 2016)).

We decided to have a look at a *pragmatic dimension* of publishing behavior and find out if there is a connection between the way the academic institutions are being financed and the way how the publications patterns are changing. In Slovakia, the ministry of education implemented a system of performance financing in the early 2000s, one of the performance factors being the publications published and registered into the national database (CREPC). All the publications are categorized and their weight contributes to the amount of money the institution receives from the state budget.

Typology of publications that is used at the universities in Slovakia was defined by Decree of ministry of education in 2008. Overall, the scale contains more than 80 types of publications. Some of them are considered to be of higher quality (e.g. A – scientific works), as they publish the results of original scholarly research, usually validated by the quality of publication resource/journal (CC journal, reviewed collection of papers etc.).

In the course of previous 10 years, the ministry changed the weight of certain categories of publications several times. We supposed these changes had an impact on authors' decision to publish their research results in certain types of papers. We concentrated predominantly on monographs, as a type of publication that is important mostly for humanities and social sciences, and CC (Current Contents) journals papers, as the category that is typical for sciences and medicine and that became a highly valued item on the list of publications. We also tried to analyze the publication language category as we had indications that the highly valued aspect of "being published abroad" is getting more and more discredited by the fact that monographs are published with obscure publishers abroad, close to the Slovakia's border and mostly in Slovak language.

Bibliographic analysis was carried out in January/February 2017 in 2 databases. CREPČ is a national database built and maintained for the ministry of education by the CVTI SR (Centre for scientific and technical information). It contains all the publications by all the academic institutions in Slovakia in a more or less complete coverage since 2006.

The second database that was used for the bibliographic analyses was the EVIPUB database, the catalogue of publications produced by the Comenius University in Bratislava. It was used for a complementary view on a disciplinary structure of the bibliographic records. As the CREPČ database does not include the aspect/field of "research area" in all of its retrospective (it was added only some 4 years ago), the possibility to stratify the production of the Comenius University by individual faculties gives us a chance to do some interdisciplinary comparisons.

The results show that there is a slight interconnection between the changes in the weight of certain types of publications (monographs, CC journals) and their "production" in the following years. Especially when it comes to categories that can be interchanged with one another within the typology of publications of the ministry of education, there is a

tendency to classify the publication as the one that "brings more money". As the system works primarily on the judgment of formal criteria, most frequently made by librarians (e.g., is publisher a Slovak or foreign entity?), it is quite often possible to do so. Of course, these conclusions are drawn just on the statistical data and it would require further qualitative analysis to give us a chance to identify the reasons behind these decisions.

Acknowledgements

This research was supported by research grant VEGA 1/0066/15 Modelovanie informačného prostredia digitálnej vedy (Modelling of information environment of digital science) and by the ITMS 26240220086 project – University Science Park, Comenius University in Bratislava.

References

Björk, B.C. & Turk, Z. (2000). A Survey of the Impact of the Internet on Scientific Publishing in Construction IT and Construction Management. In *The Journal of Information Technology in Construction (ITcon)* [online]. Rotterdam: International Council for Research and Innovation in Building and Construction, 2000, vol. 5, pp. 73-88. ISSN 1874-4753. Retrieved February 26, 2017 from: http://www.itcon.org/2000/5/paper.htm.

Kyvik, S. (2003). Changing trends in publishing behavior among university faculty, 1980-2000. In *Scientometrics* [online]. Budapest: Akademiai Kiado Rt. September 2003, vol. 58, iss. 1, pp. 35-48. ISSN 1588-286. Retrieved February 26, 2017 from: http://www.springerlink.com/content/t779563774467186/.

Rockwell, G. & Siemens, L. (2000). *The Credibility of Electronic Publishing : Report on Responses to the Questionnaire* [online]. Nanaimo: Malaspina University. Retrieved February 26, 2017 from: http://web.mala.bc.ca/hssfc/Final/QuestionnaireR.htm

Steinerová, J., Šušol, J. & Grešková, M. (2006). Information behavior in relevance judgements. In *Využívanie informácií v informačnej spoločnosti.* Bratislava: Centrum VTI SR, pp. 29-40. ISBN 80-85165-92-9.

Steinerová, J. (2016). Information behavior of researchers: contexts of digital scholarship. In *WIKT & DaZ 2016 : proceedings in informatics and information technologies.* (pp. 160-165) Bratislava: STU.

Šušol, J. (2004). Publikačné preferencie autorov vo vzťahu k tradičným a elektronickým zdrojom - situácia v oblasti vedeckej komunikácie. In *Knižničná a informačná veda 20 = Library and information science 20*. Bratislava: Univerzita Komenského, 2004, pp. 37-57. ISBN 80-223-1993-7.

Verleyssen, F. & Engels, T. (2015). Clustering of authors through the analysis of their publication patterns. In *RESSH 2015*. Retrieved February 26, 2017 from: http://www.evalhum.eu/RESSH2015/presentations/VerleyssenEngels.pdf.

Quality criteria and research obstacles in the SSH in Macedonia

Michael Ochsner and Miso Dokmanovic

Research evaluation in the social sciences and humanities (SSH) is a delicate issue. On the one hand, scholars in the SSH are often very sceptical towards evaluation and on the other hand evaluation methods reflect research practices in the science, technology, engineering and Medicine (STEM) disciplines but are not adapted to the processes of knowledge production in the SSH (see e.g. Nederhof, 2006). Furthermore, research evaluation is always related to specific policy or governance goals (Loprieno, Werlen, Hasgall & Bregy, 2016), which is rarely reflected in the evaluation processes. Rather, the same indicators are used in very different situations.

In this work-in-progress presentation we tackle these two issues, i.e. adequate criteria for SSH research evaluation and the contextual situation of evaluation. While there exist studies on SSH scholars' quality criteria for some research-intensive countries (see e.g. Hemlin, 1993; Hug, Ochsner & Daniel, 2013; Oancea & Furlong, 2007), there are no comparable studies for less research-intensive countries. However, the functions and tasks of the universities in these countries might differ and hence the priority of quality criteria can be different. Furthermore, research conditions matter in achieving research quality, especially so in countries with scarce research infrastructure and funding. To investigate quality criteria in a less research-intensive country, we focus on the research situation in Macedonia. Macedonia is a young state emerging from former Yugoslavia and only a small amount of money is spent on research (0.4% of GDP compared to 2.0% in the European Union in 2013, see World Development Indicators, 2017). Macedonian scholars, especially in the social sciences and humanities, face a lot of insecurities: Amounts of funding promised by the government can be cut back after projects have already started, there is a lack of transparency on government spending, institutions risk not receiving any funding over a certain period etc. (Cvetkovic, 2013; Petresky, 2013). Funding is also scarce and it is difficult to finance up-to-date software or access to state-of-the-art literature (Dokmanovic & Gicevska, 2013). At the same time, researchers face a rigid evaluation system since a new Law of Higher Education has been adopted in 2011 and they have to publish in journals indexed in Web of Science to get tenure (Petresky, 2013). Therefore, research quality is especially linked to research conditions in Macedonia.

In our presentation, we will present the first results of a survey conducted in Macedonia on quality criteria for SSH research and obstacles that Macedonian researchers face when doing research in a restrictive environment. For the first part, we started from the catalogue of quality criteria for humanities research (Hug, Ochsner & Daniel, 2013) and adapted the catalogue to cover the social sciences and the situation in Macedonia. For the second part, we chose the pertinent questions from a Macedonian study on challenges for young researchers in the social sciences (Dokmanovic & Gicevska, 2013) that cover problems Macedonian researchers face when conducting research. The survey will be fielded among all Macedonian SSH scholars between April and June 2017.

Using the results of the survey, we will identify quality criteria that reach consensus among Macedonian SSH scholars. We will compare these results to the results obtained by the similar survey conducted in Switzerland using bootstrapped stability intervals (see Schneider & van Leeuwen, 2014). We will furthermore identify the most pertinent obstacles for conducting research and link them to the quality criteria on which the obstacles might have an effect.

In our presentation, we will focus on commonalities and differences in the notions of quality of Macedonian and Swiss scholars, two countries that face different problems in research policy and have very different approaches to research evaluation. We will then reflect on the implications of a country's situation in the research landscape on evaluation practices in the SSH.

References

Cvetkovik, I. (2013). Macedonian scientific reality: challenges and perspectives. In Indzevska S and Dimova Mancevska A (eds.) Research in social sciences in Macedonia. State-of-affairs, challenges and recommendations for public policy improvements (pp. 148-152). Skopje: Foundation Open Society – Macedonia.

Dokmanovic M and Gicevska S (2013). Perspectives and challenges for young researchers in the field of social sciences and humanities in the Republic of Macedonia. In Indzevska S and Dimova Mancevska A (eds.) Research in social sciences in Macedonia. State-of-affairs, challenges and recommendations for public policy improvements (pp. 73-102). Skopje: Foundation Open Society – Macedonia.

Hemlin, S. (1993). Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics, 27*(1): 3–18.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation* 22(5), pp. 369–383. http://doi.org/10.1093/reseval/rvt008

Loprieno A, Werlen R, Hasgall A and Bregy J (2016). The 'Mesurer les Performances de la Recherche' Project of the Rectors' Conference of the Swiss Universities (CRUS) and Its Further Development. In: Ochsner M, Hug S E and Daniel H-D (eds.) *Research Assessment in the Humanities. Towards Criteria and Procedures* (pp. 13–21). Springer International Publishing: Cham.

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

Oancea, A., & Furlong, J. (2007). Expressions of excellence and the assessment of applied and practice-based research. *Research Papers in Education, 22*(2): 119–137.

Petrevski, M. (2013). From economic demagogy to economic research in Macedonia: Sisyphean Task? In Indzevska S and Dimova Mancevska A (eds.) Research in social sciences in Macedonia. State-of-affairs, challenges and recommendations for public policy improvements (pp. 9-40). Skopje: Foundation Open Society – Macedonia.

Schneider J W and van Leeuwen, T N (2014). Analysing robustness and uncertainty levels of bibliometric performance statistics supporting science policy. A case study evaluating Danish postdoctoral funding. *Research Evaluation* 23(4), pp. 285-297.

The World Bank. (2017). World Development Indicators. Research and Development Expenditure (% of GDP), GB.XPD.RSDV.GD.ZS. New York, NY: The World Bank Group.

National research assessments

Changing publication practices in SSH: Evidence from two consecutive national research assessment exercises (VQR 1, 2004-10; VQR 2, 2011-14)

Antonio Ferrara and Marco Malgarini

Performance based funding systems have a significant impact on the actual funding of universities and other research institutions; in this way, they may also contributing to modify habits, methods and practices of publication, especially in humanities and social sciences (HSS). The aim of our paper is to shed some light on this issue, considering publication practices of Italian researchers in the last ten years as they emerge from two consecutive evaluation exercise (that is, VQR 1, 2004-2010, and VQR 2, 2011-14). The scientific disciplines taken into consideration include Anthropology, Architecture, Archeology and Classics, Art History, Geography, History, Economics and Statistics, Education Science, Library Science, Linguistics, Philology and Literature, Philosophy, Law, Sociology and Political Science.

In HSS books, book chapters and monographs represent a significant, sometimes dominant, share of scientific production, while journal articles are less central, and national languages are widely used (Finkenstaedt, 1990; Hammarfelt, 2012). SSH researchers, having generally a more pronounced 'national' orientation than STEM ones, are thus more likely to use their native language for academic purposes (Bolton and Kuteeva 2012, Kuteeva and Airey 2013). However, research assessment processes are alleged to use evaluation modalities better suited to the specificities of natural sciences; so, they might be pushing HSS researchers towards modes of dissemination (if not production) of knowledge closer to those usually considered as more typical of STEM. If this was the case, there would be reason to expect a shift away from 'national literatures' (as defined by Hicks 2004) – all the more so because, in the Italian case, encouraging internationalization has been an explicit goal of research evaluation. The latter thus could (and perhaps should, to some extent) have had an impact on the 'trade-off between local and international publication' (Flowerdew and Li 2009) among Italian HSS academics, driving an increase in the quota of foreign-language publications – especially of English-language ones, accessible to a wider readership and often published in international (and/or indexed) journals. The larger issue, undergirding to an extent all of those mentioned below, is thus whether research assessment processes are influencing publication practices in the HSS. Of course, it is difficult to disentangle the effects of research evaluation from those of wider, world-wide trends towards the increasing use of English in academic publication. Those trends are, nonetheless, affected by the globalization not only of research communication, but of research assessment as well – and by policies rewarding English-language publications rather than national-language ones (Li and Flowerdew 2009, Li 2014), even where researchers are encouraged to use their mother tongue to communicate their results (López-Navarro et al. 2015). Thus, the rich troves of data provided by two consecutive, large-scale national research assessment exercises can, at the very least, provide us with useful information on whether, and to what extent, Italian SSH research is (or not) part of those trends.

In the paper, we will first try to understand if, and how, the quota of foreign-language publications has changed over the years, and is there any relationship between the language of publication and research quality (as measured through the scores assigned by VQR reviewers to evaluated publications). Secondly, we will examine how the distribution of scholarly works between different "literary genres" (that is, monographs, book chapters, journal articles, etc.) has changed over the years, and will strive to unearth if there is any interrelationship between the literary genres and the quality of the assessed works. To this end we will focus on a subset of journals – that is, those identified as 'A-class' for other purposes (namely, those of the National Scientific Habilitation) – in order to discern whether they account (or not) for a greater share of publications in VQR 2 than in VQR 1. We will also try to ascertain if the articles published on those journal keep getting consistently better evaluations than those appearing in other journals (as argued in Ferrara and Bonaccorsi 2016 and Bonaccorsi, Cicero, Ferrara and Malgarini 2015). We will also address the somewhat related question of the quota of indexed (in WoS and/or Scopus) journal articles; as in some areas such publications *de facto* guaranteed a positive evaluation, one might expect that the share of such publications to have increased over the years, partly as an effect of the enactment of research assessment processes. Lastly, we will address the question of co-authorship: this particular publication practice tends to be much more widespread in STEM than in the SSH, yet data from two consecutive national research assessment exercises can be used to discern both the relative dimensions of the gap. They can tell us as well as if, in time, co-authorship is becoming more pervasive also in the SSH.

In the end, we expect our paper to contribute to the ongoing debates on the role of evaluation in knowledge production processes and the effects of evaluation and performance-based funding on SSH research and researchers. To some extent, it will also contribute to debates concerning the interaction between research evaluation on one hand and the evolving role of the English and of national languages in academic publishing in the SSH on the other.

References

Bonaccorsi A, Cicero T, Ferrara A. and Malgarini M., Journal ratings as predictors of articles quality in Arts, Humanities and Social Sciences: an analysis based on the Italian Research Evaluation Exercise [version 1; referees: 3 approved]. F1000Research 2015, 4:196 (doi: 10.12688/f1000research.6478.

Bolton, K. and Kuteeva, M. (2012). English as an Academic Language at a Swedish University: Parallel Language use and the 'Threat of English', Journal of Multilingual and Multicultural Development, 33/5: 429–47

Ferrara A., Bonaccorsi A. (2016). How robust is journal rating in Humanities and Social Sciences? Evidence from a large scale multi-method exercise, Research Evaluation, 25(3), 279-291

Finkenstaedt, T. (1990). Measuring research performance in the humanities. Scientometrics, 19(5/6), 409-417.

Hammarfelt, B. (2012). Harvesting footnotes in a rural field: citation patterns in Swedish literary studies. Journal of Documentation, 68(4), 536–558

Flowerdew, J., & Li, Y. (2009). English or Chinese? The trade-off between local and international publication among Chinese academics in the humanities and social sciences. Journal of Second Language Writing, 18, 1-16.

Hicks, D. (2004) The four literatures of social science. In: Moed, H.F., Glänzel W., Schmoch U. (eds.) Handbook of Quantitative Science and Technology Research. Dordrecht, Kluwer Academic Press, 473–496.

Kuteeva, M., & Airey, J. (2013). Disciplinary differences in the use of English in higher education: reflections on recent language policy developments. Higher Education, 67(5), 533-549.

Li, Y., & Flowerdew, J. (2009). International engagement versus local commitment: Hong Kong academics in the humanities and social sciences writing for publication. Journal of English for Academic Purposes, 8, 279–293

Li, Y. (2014). Seeking entry to the North American market: Chinese management academics publishing internationally. Journal of English for Academic Purposes, 13, 41-52.

López-Navarro, I., Moreno, A. I., Quintanilla, M. Á., & Rey-Rocha, J. (2015). Why do I publish research articles in English instead of my own language? Differences in Spanish researchers' motivations across scientific domains. Scientometrics, 103(3), 939-976.

Staging research impact. How academics write and talk about the wider impact of their research in the context of REF

Marta Natalia Wróblewska

The introduction of 'impact' as an element of assessing academic work is a major change in the way scientists construct and conceptualize the value of research. Though different systems of evaluation of impact play a role in many national and supra-national contexts (eg. Netherlands, Sweden, Australia, EU), the most robust system implemented to date is the British REF 2014 with its Impact Agenda.

My research focuses on the effect of this development on the way academics conceive of the value of research and their own role in society. I take a social constructivist approach in assuming that 'daily activities of working scientists lead to the construction of scientific facts' (Latour & Woolgar, 1986, p. 40) and condition the understanding of academic values. I use discourse analysis to analyse empirical textual data and reflect on the relation between everyday practices and the emergence of new concepts in academia.

In my talk I will present findings from the analysis of two corpora: case studies submitted by British linguists to REF 2014 (no ≈ 100) and interviews (n ≈20) with their authors. When working with the first corpus, I use pragmatic methods, notably genre analysis (Swales, 1990), to describe the new emergent genre of academic writing – impact case study. I reflect on the rules of this genre and attempt to answer the question 'what is the purpose of this type of writing' i.e. what vision of research do such documents present?

When working with the second corpus (interviews), I reflect on how academics position themselves (Angermuller, 2013b; Bamberg, 1997) towards 'impact'. My data shows that academics' approaches are far from being simply positive or negative (Chubb, Watermeyer, & Wakeling, 2016) or evolving from one to the other in time ('acceptance'). Rather, researchers have complex (sometimes even contradictory) views and attitudes towards 'impact' in the same time. I conceptualise this differentiation building on Goffman's notion of 'stage' (Goffman, 1956) and Angermuller's distinction between scientific and institutional spheres in academic careers (Angermuller, 2013b, 2016). I argue that academic's attitude and valuation of impact is a complex negotiation between views they present on the scientific, institutional and personal 'stage'.

I conclude with a more general reflection on the differences in the way research is presented in the context of evaluation (on the 'institutional stage') and on the scientific and personal stages. Finally I pose an open question: what effect does the use of rhetorical moves such as boasting, presenting linear narratives and excessive focus on numbers (Angermuller, 2013a) have on the wider public's vision of science.

I believe the topic of my research corresponds with at least two themes of your conference, namely: 'Evaluation of societal relevance of SSH research' and 'Effects of evaluation and performance-based funding on SSH research and researchers', and therefore would be a valuable contribution to the event.

Keywords: Research Excellence Framework, Societal Impact of Research, Impact Agenda, Discourse Analysis, Qualitative Research, REF 2014

References

Angermuller, J. (2013a). Discours académique et gouvernementalité entrepreneuriale. Des textes aux chiffres. In M. Temmar, J. Angermuller, F. Lebaron (Ed.), *Les discours sur l'économie*. Paris: PUF.

Angermuller, J. (2013b). How to become an academic philosopher. Academic discourse as a multileveled positioning practice. *Sociología Histórica, 2*, 263-289.

Angermuller, J. (2016). The discursive constitution of academic job markets. Academic categorisation systems and valuation practices in France, Germany, Great Britain and the U.S. *[under review]*.

Bamberg, M. (1997). Positioning Between Structure and Performance *Journal of Narrative and Life History, 7*(1-4), 335-342.

Chubb, J., Watermeyer, R., & Wakeling, P. (2016). Fear and loathing in the Academy?:The role of emotion in response to an impact agenda in the UK and Australia (Forthcoming). *Higher Education Research and Development*.

Goffman, E. (1956). *The presentation of self in everyday life*. Edinburgh,: University of Edinburgh, Social Sciences Research Centre.

Latour, B., & Woolgar, S. (1986). *Laboratory life : the construction of scientific facts*. Princeton, N.J.: Princeton University Press.

Swales, J. M. (1990). *Genre analysis : English in academic and research settings*. Cambridge [England]; New York: Cambridge University Press.

Research groups in SSH – A typological analysis of 228 groups submitted to national research evaluations in Norway

Jon Holm and Heidi Dybesland

The organisation of research groups has been a prominent subject of scholarly study at least since the 1970ies (Andrews 1979, cited in Kyvik et al. 2015). Most of the studies are based on data from the STEM fields. There is thus a risk that the models we use for describing and assessing research organisation are not well adapted to the social sciences and the humanities. Studies of research groups in SSH carried out in Norway in the later years (Kyvik, S. 1998 and Kyvik et al. 2015, Gulbrandsen 2016) indicate that the share of researchers in SSH that perform research as a part of a group is increasing: In a survey among research personnel in the Norwegian HE-sector carried out in 2013, the share of staff reporting that they perform some or an important part of their research in a research group was 61% and 47% within the social sciences and the humanities respectively. At the oldest research universities the share was even higher, around 70% within the social sciences and over 50% within the humanities (see Kyvik et al 2015 p.16-17).

In this paper we will provide the first systematic presentation of a unique empirical material on research group organisation collected in the context of national research evaluations in Norway. In two recent evaluations of the humanities (2015-2107) and the social sciences (2016-2018), the Research council of Norway has introduced research groups as a unit of assessment. The purpose of the evaluation of research groups is to identify groups within SSH in Norway that are internationally leading or that have the potential to become a world leader in their field. This purpose corresponds to a political ambition in Norway not only to raise average quality but also to develop and sustain some world-leading groups (Meld. St. 7 (2014–2015).

The use of research group organisation as a policy-instrument to promote excellence is still controversial in some of the SSH research communities. When the Research Council of Norway (RCN) announced its intention to include research groups in the next national evaluation of the humanities, this immediately spurred a debate on the appropriateness of the research group model within the humanities. Prominent scholars criticised RCN for not taking into consideration that "the large majority of researchers in the Humanities work and publish alone" (Førland 2015).

The polarisation of the debate is in itself interesting but also problematic. Judgements on the appropriateness of research group organisations within SSH seems to be deeply rooted in the self-perception of scholars as autonomous agents that are free to seek or to decline collaboration with other autonomous individual scholars without the more permanent mutual obligations and coordination of activities towards a common goal that is characteristic of a more formal research group organisation.

Within the social sciences, research group organisation seems less controversial and is becoming increasingly important in the HE-sector (see Kyvik et al 2015 p. 9). At the social science research institutes, organisation in research groups is often the norm. The institutes

often engage in cross-disciplinary research, and are more dependent on external funding in terms of larger research projects. Research groups are often effective organisational responses to these requirements.

There is a tendency that research group organisation is becoming a more normal feature of SSH policies and practices. At the policy level, SSH is included in group oriented funding schemes at the research council, like the Centres of Excellence scheme (and a new scheme targeting "dynamic research groups" at a smaller scale (FRIPRO Toppforsk). The most research intensive universities have also put in place measures to support research groups within SSH. According to Kyvik et al (2015) these measures are a response to a variety of institutional policy aims, like strengthening of internal cooperation, support of research leadership, securing a good environment for researcher training at PhD and postdoc levels, creating hubs for national and international collaboration as well as stimulating project development for external funding. The facilitation of research cooperation across fields and disciplines is a recurrent topic in the organisation of the HE-sector. Organisational frameworks like research groups may also be understood as part of this trend. The political and institutional policies for excellence with an associated focus on the development of more formalised research groups make it imperative to understand more precisely how research organisation in SSH is related to research quality and international competiveness.

The data material on research groups submitted to the RCN gives us a valuable empirical basis for the study of an emerging practices of research group formation also in SSH. Even if we need to be cautious of observer effects, i.e. the influence of the definition of the research group as a unit of evaluation on how the groups' submitted to for evaluation describe their organisation in the self-assessments, the variation in these descriptions among the 228 groups could still be taken as a token of variation in research culture and policies in different disciplines and institutions. In this paper we will analyse this variation in order to suggest a typology of SSH research groups and compare it to the typologies proposed by the standards literature on research groups, which is mainly based on data from STEM fields. In this way we hope to contribute to a better understanding of the specific characteristics of research organisation within SSH as distinct from STEM.

The 228 submitted research groups submitted to the evaluation exercise will be analysed according to quantitative characteristics – like funding, group size, members' age, position, gender and affiliation, co-publication, and research quality as assessed by evaluation panels – and qualitative characteristics – like organisation and leadership, expression of goals and the group's engagement in teaching, research training, knowledge exchange, networking and research impact. The data material will also enable us to assess whether these characteristics vary between the humanities and the social sciences and across different types of institutions. Based on such characteristics we will propose a typology that could be refined through further research and hopefully contribute to a better adaptation to SSH of the definition of research groups as units of assessment in future evaluations.

References

Andrews, F.M. (red.) (1979). Scientific Productivity. The Effectiveness of Research Groups in Six Countries. Cambridge/Paris: Cambridge University Press/UNESCO.

Førland, TE (2015). "Et vrengebilde av humanistisk forskning" [A distorted image of humanities research – Newspaper chronical] in *Morgenbladet* 6 November 2015. https://morgenbladet.no/ideer/2015/11/et-vrengebilde-av-humanistisk-forskning

Gulbrandsen, M (2016). The humanities in Norway: research, research organisation and external engagement. Oslo: NIFU Report 36-2016 http://www.nifu.no/publications/1406295/.

Kyvik, S (1998). Kritisk masse – om forskningsmiljøers størrelse, produktivitet og kvalitet. [Critical mass – on the size, productivity and quality of research milieus] Oslo: NIFU skriftserie 8/98.

Kyvik S et al. (2015). Forskergrupper i universitets- og høgskolesektoren [Research groups in the HEIs]. Oslo: NIFU working paper 2-2015 http://www.nifu.no/publications/1234418/.

Meld. St. 7 (2014–2015). Langtidsplan for forskning og høyere utdanning 2015–2024

The Quest for 'Research Quality' in the Humanities and Social Sciences: A Japanese perspective on the Research Excellence Framework (REF)

Eriko Amano, Ayako Fujieda, Natsuko Inaishi, Toshiro Kamiya, Akiko Morishita,Asa Nakano, Yoshimi Osawa, and Yu Sasaki

On June 8th 2015, the Ministry of Education, Culture, Sports, Science and Technology of Japan issued what was quickly understood as an unprecedented Notice[2] to all national universities and higher education institutions in the country. The Notice strongly advocated that the universities were required to reorganize themselves because of a steep decline in the demographic that was available for university education in Japan. The Notice also bluntly recommended that the Humanities and Social Sciences (HSS) departments take active steps to either shut down or drastically re-orient themselves towards more 'practical' research that reflected the 'actual' needs of society.

While the subsequent storm of criticism[3] against the Notice led the Ministry to substantially retreat from its earlier positions, the issue of research quality assessment in the HSS has nonetheless been strongly foregrounded. In other words, it was now felt that bringing in accountability into research, assessing research output and enabling the HSS in Japan to be an active part of a global academic dialogue had acquired a new urgency. While the academic associations were tasked to design protocols for defining research standards, these associations have thus far failed to arrive at a consensus on quality norms. It is in this context that newly organized research administrators along with various stakeholders in the university system in Japan had begun to actively debate distinct strengths and weaknesses in the HSS. Part of such attempts is a debate within Kyoto University Research Administration Office (KURA) on the relevance and efficacy of Research Excellence Framework (REF) for the HSS in the national university system in Japan.

The REF exercise, an evaluation effort for assessing British higher education institutions, was instituted in 2014 to succeed the earlier Research Assessment Exercise (RAE) that has periodically taken place since 1986. The RAE and REF are undoubtedly amongst the most debated and widely discussed research assessment exercises in recent times. This paper will outline and explore the various assessments, critiques and questions that KURA raised as part of its internal discussions on the REF. We point out here that three main overarching frameworks — also marked by several topical and thematic sub-debates — emerged in assessing the relevance of the REF for the Japanese HSS context:

a) HSS as Competitive or Cooperative endeavour?

Historically the HSS tradition in Japan has been inseparably linked to the modernization project and the HSS was in a way a tool to produce citizens/leaders who were then

expected to become part of modern nation building. HSS pursuits are 'non-competitive' academic fields, in which the ability to question and ethics and values (such as democracy) are central to the research and teaching efforts. In effect, HSS is better understood as a 'cooperative exercise' that inherently premises diversity within and across disciplines even as the challenge of quality and standard setting remains. Given such a context, this paper will discuss the various debates over the cooperative versus competitive axis and notably whether the idea of 'ranking' needs to be critically examined.

b) Local Diversity against Global Homogenization

The HSS higher education context in Japan is marked by rich regional and local diversity and shaped by a strong commitment to the Japanese language. Whilst there is a compelling case to be made for translating such local/regional output and also the need to engage with global scholarly traditions in the HSS, there can be worries that any unqualified REF exercise can potentially undermine internal academic diversity by instituting in its place a homogenous academic culture which is more committed to global HSS questions. Thus, different higher education contexts that hold uniquely for Japan need to be addressed with a far more nuanced evaluation criteria.

c) Inter-disciplinary, transdisciplinary and multi-disciplinary

Japan's unique demographic, environmental and technological challenges have also increased emphasis on inter-disciplinary, trans-disciplinary and multi-disciplinary programs. Can the REF exercise speak to this unique requirement has also been part of our discussion, in which the HSS cannot be seen to be distinct research domain. Rather the HSS needs to be seen as a critical bridge building exercise with the fundamental sciences and the technical studies such as engineering, robotics, artificial intelligence etc.

While KURA conducted several levels of discussions under these three broad frameworks to judge the relevance of the REF for the Japanese HSS context, there were also a range of logistical and ' known criticism' within the UK of the REF that were pointed out as well. In particular, the fact that the REF exercise involved a great deal of financial investment not only to institute it but also to run it on a regular basis. Monitoring quality, in other words, is time- and cost-consuming and raises the issue of opportunity costs as well. It was also pointed out that academic monographs were being replaced by peer reviewed journal articles which were considered more 'REF-able' and this had impacts on long term research projects.

By putting forward Japanese perspectives on the REF, this paper will also highlight the status and mood of the HSS situation in Japan. The effort is intended to not only help us understand the impacts of globalization on the HSS context in Japan but to also suggest that dialogues, adjustments and rejection with regard to global ranking systems and research quality standards will be part of the dynamic for change in the academic community in the years to come.

References

Government of Japan (2015). *A Review of the Organization and Operation of the National University Corporations and Other Higher Educational Institutions* (in Japanese). Ministry of Education, Culture, Sports, Science and Technology. Retrieved February 27, 2017 from: http://www.mext.go.jp/b_menu/shingi/chousa/koutou/062/gijiroku/__icsFiles/afieldfile/2015/06/16/1358924_3_1.pdf.

Science Council of Japan (2015). *On the Future Direction of the University: In Relation to the Departments/Graduate Schools of Teacher Training and Humanities and Social Sciences*. Retrieved February 27, 2017 from: http://www.scj.go.jp/en/pdf/kohyo-23-kanji-1e.pdf.

2 Website: http://www.mext.go.jp/b_menu/shingi/chousa/koutou/062/gijiroku/__icsFiles/afieldfile/2015/06/16/1358924_3_1.pdf

3 For instance, website; http://www.scj.go.jp/en/pdf/kohyo-23-kanji-1e.pdf

Peer review

Game theory, crowdsourcing and the peer review system

Imre Lendák and Karolina Lendak-Kabok

Introduction

Journal editors and conference organizers often face challenges when looking for suitable reviewers to review the papers submitted to their journals or conferences. The shortage of timely and high quality reviews leads to delays in the review process. Authors all too often have to wait for the reviews for extended periods of time and sometimes they receive only one review in the end. The reviews might be submitted by non-experts, who provide limited feedback without sufficient detail. Some reviews address only the formatting and language issues, as the reviewers do not risk going into criticizing the content which lies outside their core domains of expertise. Apart from authors, editors are also dissatisfied, because they often find themselves lacking reviews in moments when the paper notification deadlines are upon them. Desperate editors and/or conference organizers often complete a significant number of reviews by themselves or assign them to researchers or teaching assistants in their groups.

This leads us to the conclusion that the peer review system, which consists of conferences/journals, editors, authors, papers and reviewers, still relies on volunteer reviewers, which are hard to find. Leading researchers and experts usually do not have the necessary time to devote to volunteered reviews. Younger researchers are more easily convinced to review, but their lack of experience might affect the quality of reviews they submit. Some domain experts even claim that the peer review system is broken (Baldwin, 2014; Benos et al, 2007). Some researchers claim that the complete system is biased (Lee et al, 2012), or that the reviewers are specifically biased toward their social groups to whom they give better reviews (Wenneras & Wold, 2001; Travis & Collins, 1991). Others argue that the only realistic way forward is switching to a for-profit peer review system, in which reviewers are paid for their services by journals. Many oppose this direction, as it might lead to a situation in which the rich publish and the less well-off 'perish'.

The goal of this paper is to present a simulation environment capable to investigate various peer review scenarios involving non-financial reviewer motivation techniques in peer review systems involving thousands of reviewers. In this context, the reviewers form a crowd which solves reviewing tasks, thereby creating a crowdsourced system. The authors analyzed various scenarios in the simulation environment in which the journals and/or editors use gamification techniques to motivate their prospective reviewers, e.g. badges assigned after a certain number of reviews, leaderboards in which reviewers compete with other reviewers in their specific research domain. The results achieved were compared to simulations of the traditional review system consisting of a limited number of high quality journals/conferences, as well as to the modern scientific peer review system which consists of a plethora of journals and conferences of varying quality. The ultimate goal of the authors is to propose such modifications to the peer review system which would make it more transparent and efficient on one hand, and more fun via the introduction of gamification techniques on the other hand.

Problem definition

The following are the key elements of the peer review system:
1. journals and conferences, i.e. the venues of publication
2. editors and conference program chairs who among other things, own the review process
3. authors who submit papers, and
4. reviewers who review the papers and thereby provide advice to the editors.

The current peer review system relies on volunteer reviewers, who are not paid for their reviewing activity. They are usually motivated by intrinsic altruism, minimum review effort enforced by institutions or journals/conferences or plain old scientific curiosity. Unfortunately, with the constant increase in the number of conferences and journals, it is becoming harder to find suitable reviewers, who provide detailed and useful reviews in a timely fashion.

Solution

The authors modeled the peer review system (PRS) as a multi-agent system in which the reviewers are autonomous agents, who either accept or decline reviews based on a complex decision formula, calculated from the perceived quality of paper to be reviewed, its similarity to the research done by the reviewer, the amount of free time each reviewer has, journal or conference prestige (e.g. more likely to accept a review in a leading journal).

The model takes into consideration the following additional incentive mechanisms borrowed from the field of gamification:
1. a (theoretic and) globally centralized and available **point system** which consists of every researchers and expert who submits reviews – this system records each review submitted in the world and reward reviewers for reviewing activities, e.g. 5 points for a submitted review at a leading conference within a certain scientific domain,
2. **leaderboards** of most active reviewers in a specific scientific sub-domain, e.g. information security, citizen science, life cycle analysis,
3. special **tiers** within scientific sub-domains assigned to reviewers after obtaining a certain number of points, e.g. fledgling reviewer with 500 points, master reviewer with 1500 points, etc.
4. **badges and achievements** received after a certain number or certain type of reviews submitted,
5. **mini-games** which are specific, time-limited challenges presented to reviewers, e.g. complete two reviews at each of three specific conferences to receive a specific reward.

The above listed additional incentive mechanisms would rely on a globally available reviewer database, in which each reviewer and each review would be recorded. Such system would have certain upkeep costs in computation power and storage space, as well as maintenance services.

Results

The authors developed a multi-agent simulation environment in the Java programming language and by relying on the MASON simulation environment (Luke et al, 2005). The resulting simulation environment was used to simulate the following three peer review scenarios:

1. the traditional peer review system (PRS) with a small number of high quality journals with high acceptance rates,
2. the current PRS with numerous journals of varying quality and medium-level acceptance rates and without additional motivation techniques from the gamification field,
3. a future PRS with numerous journals of varying quality and medium level acceptance rates, but with additional motivation techniques from the gamification field.

The PRS performance was measured via the number of accepted and rejected papers, as well as the number of papers with zero reviews, average review time (expressed in weeks) and journal editor/conference chair load. The organizers' load was expressed via the number of (potential) reviewers the organizers had to contact to receive one detailed review.

Additional constants used during the simulations were the average quality of papers received (e.g. 7 points out of 10 for high quality journals), reviewer motivation to accept paper reviews (e.g. 2 out of 10), level of reviewer load (e.g. 10% of the workweek devoted to reviews) and the paper acceptance rates (e.g. 30%). All simulations lasted 5 years with a one-week resolution. This meant that in each week of the simulation the author 'agents' submitted papers within their three areas of expertise with a certain predefined probability. The probability of reviewers' review acceptance was calculated in simulation-time.

Experiment 1: High quality journals

In the first experiment the simulation environment was configured with 10 high-quality journals, 1000 scientists who might submit papers, 10% of the scientists were accepting reviews, and 3/10 average paper quality. This scenario was used to simulate the peer review system as it might have looked like in the 1970s and 1980s, as well as a baseline scenario aimed at tuning the default parameters of the simulation environment. The simulation results were aligned with the authors' expectations, as in the course of the 5-year-long simulation period only 310 papers were accepted, 792 rejected (mostly due to the low paper quality setting) and zero papers failed to receive any reviews. The average review time was 5 weeks and the editor/organizer load was 1:2, i.e. they had to contact two potential reviewers to get one review. The results of this experiment are presented in row one of Table 1.

Table 1. Overview of simulation results.

Simulation	Paper Quality	Reviews [Accepted / Rejected (No review)]	Review time [week]	Editor load
Few strong journals	3 / 10	310 / 792 (0)	5	1:2
Many journals	3 / 10	4976 / 8928 (2154)	7	1:30
Many journals + gamification	3 / 10	5834 / 9004 (750)	7	1:12

Experiment 2: Many journals

In the second experiment the authors simulated the current state of the peer review system in the 2010s, in which there are many journals (100) of varying quality and significantly more researchers (10000), both authors and reviewers (also 10% of the total). The reviewers' motivation to accept paper reviews was configured to be lower, as well as the available time they had for reviews.

The results of this experiment are shown in row two of Table 1. The acceptance rates were higher (see the accepted/rejected ratio in column three) and the review period was extended to seven weeks. The most notable change in this simulation was the editor load, which increased to 1:30, while in experiment 1 it was only 1:2. This meant that the editors had to contact 30 potential reviewers on average to get one review. Unfortunately, this simulation result was aligned with the personal editorial experience of the authors.

Additionally, due to the increased total number of papers submitted, lower motivation and high editor overhead, some papers received zero reviews in this scenario, while all papers were reviewed in experiment one.

Experiment 3: Many journals with gamification

In the third experiment the authors kept the majority of settings configured for the second scenario (100 journals, 10000 researchers, low motivation to review, no time). The most notable difference compared to the previous experiment was the introduction of two additional gamified motivation techniques, namely a reviewer leaderboard and badges received after a certain number of reviews completed. These motivation techniques were used to slightly increase the otherwise low motivation to review. Not surprisingly, the total amount of reviews received was measurably increased (compare rows 2 and 3 in Table 1), the number of papers with zero reviews significantly lowered and the editor load was also lowered by ~60% (1:12 compared to 1:30).

Conclusion

The authors developed a multi-agent simulation environment for investigating various scenarios in the peer review system (PRS). The simulation environment was used to simulate the traditional PRS with only a couple of high quality journals and highly motivated reviewers eager to review papers, the modern PRS with many journals of varying quality and low reviewer motivation, as well as a theoretical, future PRS with gamification techniques involved in reviewer motivation. In the last experiment, the reviewer motivation was augmented by adding a global reviewer leaderboard within each scientific domain, as well as issuing collectible badges to active reviewers after they completed certain amounts of reviews. The simulations showed that the inclusion of such techniques can significantly increase the number of reviews received (both accepts and rejects) and lower the editor load (up to 60%) measured in the number of received reviews compared to the total number of potential reviewers contacted.

The authors intend to further extend the simulation environment with a finer-grained gamification-based motivation model, incorporate the differences between scientific domains (e.g. mathematics vs social sciences) and further tune the parameters of the simulation to obtain results as similar to the real-life peer review systems as possible.

References

Baldwin M. (2014). *Is the Peer Review Process for Scientific Papers Broken?*, Time Magazine, http://time.com/81388/is-the-peer-review-process-for-scientific-papers-broken/

Benos, D.J., Bashari, E., Chaves, J.M., Gaggar, A., Kapoor, N., LaFrance, M., Mans, R., Mayhew, D., McGowan, S., Polter, A. & Qadri, Y. (2007). *The ups and downs of peer review*. Advances in physiology education, 31(2), pp. 145-152.

Cho, K. & Schunn, C.D. (2007). *Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system*. Computers & Education, 48(3), pp. 409-426.

Chubin, D.E. & Hackett, E.J. (1990). *Peerless science: Peer review and US science policy*. Suny Press.

Editage Insights. (2013). *Peer review process and editorial decision making at journals*, http://www.editage.com/insights/peer-review-process-and-editorial-decision-making-at-journals - has good references

Goldbeck-Wood, S. (1999). *Evidence on peer review--scientific quality control or smokescreen?*. British Medical Journal, 318 (7175), pp. 44.

Harnad, S. (1996). *Implementing peer review on the net: scientific quality control in scholarly electronic journals*. Scholarly publishing: the electronics frontier. MIT Press, Cambridge, MA.

Lee, C.J., Sugimoto C.R., Zhang G. & Cronin B. (2012). *Bias in peer review.* Journal of the Association for Information Science and Technology, 64 (1), pp. 2-17.

Liu, E.Z.F., Lin, S.S., Chiu, C.H. & Yuan, S.M. (2001). *Web-based peer review: the learner as both adapter and reviewer*. IEEE Transactions on education, 44 (3), pp. 246-251.

Luke S., Cioffi-Revilla C., Panait L., Sullivan K. & Balan G. (2005). *MASON: A multi-agent simulation environment.* Simulation: Transactions of the Society for Modeling and Simulation International, 82 (7), pp. 517-527.

Mahoney, M.J. (1977). *Publication prejudices: An experimental study of confirmatory bias in the peer review system*. Cognitive therapy and research, 1(2), pp.161-175.

Travis, G.D.L. & Collins, H.M. (1991). *New light on old boys: cognitive and institutional particularism in the peer review system*. Science, Technology & Human Values, 16 (3), pp. 322-341.

van Raan, Antony. (1996). *Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises*. Scientometrics, 36.3: 397-420.

Walsh, E., Rooney, M., Appleby, L. & Wilkinson, G. (2000). *Open peer review: a randomised controlled trial*. The British Journal of Psychiatry, 176(1), pp. 47-51.

Wenneras, C. and Wold, A. (2001). *Nepotism and sexism in peer-review*. Women, Science, and Technology, pp.46-52.

In search of visibility. How do publication culture, collaboration patterns, language of publication and locality of topics influence research assessments in the SSH ?

Thed van Leeuwen, Thomas Franssen, Clara Calero Medina, Giovanni Colavizza, Ismael Rafols, and Nicolas Robinson-Garcia

In this study we focus on the aspect of the visibility of products of academic research by scholars in the SSH. We will focus on this topic by following the line of reasoning set up by Becher and Trowler, in which they apply the metaphor of the rural versus the urban, to denote the differences in the scientific landscape that is the wide diversity from the sciences to the social sciences, humanities, and law (Becher & Trowler, 2001 and Hammerfelt, 2012). In the study we will address the question to which extent the two poles in the metaphor have an effect on the organization of research evaluation across the scientific landscape. We will apply this in the context of the recently launched QRiH website in the Netherlands, which serves as a tool to help scholars in the SSH domains to design and align their research products and activities supporting self-assessment reports, finally supporting research assessment processes. The main topics of study will evolve around three main elements that often play a central role when we talk about research assessment in the SSH, namely publication culture, language of publication, and locality of topics.

In the Netherlands, a new evaluation protocol has been adopted end of 2014. This is a next version of the Standard Evaluation Protocol (SEP), which describes in broad lines the design and requirements of research assessments in Dutch research. A novelty in the most recent version of the SEP (SEP, 2014) is the increased space for societal relevance in the assessment of research, not only in the SSH, but across the research landscape. The SEP still 'separates' the academic and the societal realms, which can be seen as problematic as it creates this dichotomy, which might reflect differences in 'value-ing' particular types of work and activities over others. Some years ago, the Deans of Faculties of Humanities in the Netherlands initiated a study in order to design an instrument to assist assessment of scholarly research activities in the humanities. This more or less coincided with the arrival of the new SEP on the scene. One of the main outcomes of this study was that research and its outcomes and products should not be analyzed in a vertical way. That is, the academic and societal realms should not be considered as separate, but linked in a horizontal way, bridging the gap between the three main criteria applied in research assessment: academic and societal on outputs (of all kinds), usage of knowledge and recognition. This resulted in a website, offering a tool box for the support of research assessment in the humanities (QRiH, 2016). A similar perspective on inclusiveness in the practice of research evaluation was recently introduced by Fochler and de Rijcke, in their notion of evaluative inquiry (Fochler and de Rijcke, 2017).

In their 2001 book, titled Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines, Becher and Trawler apply the metaphor of the rural versus the urban in

describing differences in culture between academic disciplines. They apply the metaphor on a number of elements that play a role in the process of knowledge production, and the assessment of that, by looking at the ways research is organized, the degree of specialization, communication patters (as can be analyzed through publications and reference/citation patterns), and the connected reward system, to name some elements. In the study on research assessment in the SSH domains, scholars are often confronted with issues related to the differences between SSH and STEM domains. In these discussions, we feel we often have to proof that the SSH domains are indeed different as compared to STEM domains, mainly due to the dominance of the methodologies developed to measure research performance and acclaimed research excellence in the STEM domains. To test this claim of the dominance of assessment methodologies related to STEM domains, we will start from the perspective of the metaphor developed by Becher and Trowler, and critically analyze what the usage of the metaphor means for the research assessment in the SSH in comparison with STEM domains.

The operationalization of this will be done by focusing on the journal literature as covered in the Web of Science database. Although it is clear that this system has limited value for the research assessment of scholarly products and activities in the SSH domains (van Leeuwen, 2013 and van Leeuwen et al, 2016), we want to use this system to study three elements that play an important role in the assessment of research in the SSH, namely the publication culture, the language of publication, and the locality of subject topics in research in the SSH, and that might reflect that rural character intended in the metaphor by Becher and Trowler. For this we will not only work with the source publications of the WoS (which we can use for the analysis of publication language and locality of topics), but also with the full reference lists attached to this literature, also the ones that do not relate to source material in the WoS (similar to Nederhof et al, 2010). These reference lists can learn us a lot on publication and collaboration cultures (as we can analyze many types of communication not covered in WoS, such as references to books, non WoS-covered journals, grey literature), but also reference behavior taken from a somewhat different angle (what is cited, archival material, what is the 'age' of the material referred to), as that might lead to changes in views on the communication patterns in SSH domains compared to STEM domains, which might have implications for the reward systems applied to measure research quality in the SSH domains.

References

Becher, Tony, and Paul Trowler. (2001) *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. 2nd ed. Philadelphia, PA: Open University Press.

Fochler, M. & de Rijcke, S. (2017) „Implicated in the Indicator Game? An experimental Debate", Engaging Science, Technology, and Society 3 in press.

Hammarfelt, Björn. (2012) „Harvesting Footnotes in a Rural Field: Citation Patterns in Swedish Literary Studies." Journal of Documentation 68, no. 4 (July 20, 2012): 536–58. doi:10.1108/00220411211239101.

van Leeuwen, T.N. (2013) „Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship ?“ Bibliometrie - Praxis und Forschung, 1-18 (http://www.bibliometrie-pf.de/article/viewFile/173/215)

van Leeuwen, TN, E van Wijk, and PF Wouters (2016) „Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university“, Scientometrics, 106 (1), 1-16

Nederhof AJ, T.N. van Leeuwen, and AFJ van Raan, (2010) „Highly cited non-journal publications in political science, economics and psychology: a first exploration“, Scientometrics, 83 (2), 363-374

QRiH - Quality & Relevance in the Humanities. (2016) Kwaliteitsindicatoren voor de Geesteswetenschappen (KiG), https://www.qrih.nl/nl/

Robinson-Garcia, van Leeuwen, Rafols, „Altmetrics as context indicators for evaluation: From measuring societal impact to mapping social engagement.“ *Submitted to Science and Public Policy*

SEP - Standard Evaluation Protocol for the Netherlands, (2014)VSNU, http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf

Quality Criteria for Ex-Ante Evaluation of Research Proposals from Young Humanities Scholars

Michael Ochsner and Sven E. Hug

Recently, the awarding of research grants gained in importance for research careers. With the shift to the notion of excellence in higher education institutions, the demand for excellent researchers has become more explicit. To provide excellent scientific work, we need excellent scholars (van Arensbergen, van der Weijden, & van den Besselaar, 2014a). This demand puts the focus on the selection of talent. Therefore, governments and universities put more attention on talent (selection) processes and policies and increase the support for early career researchers. The notion of excellence includes competition and such policies and sources of support often base on competitive research funding. At the same time (and directly linked to this development), work contracts for young researchers shift towards temporary employment (van Arensbergen, van der Weijden, & van den Besselaar, 2014b) Thus, the acquisition of research grants gains importance for young scholars (van Arensbergen et al., 2014a), especially because the successful acquisition of grants is considered a sign of talent (van Arensbergen et al., 2014a; 2014b). Even though research grants were not very important in SSH disciplines (especially in the humanities) until recently, the focus on talent selection and the temporary employment applies also to SSH early career scholars, thus making the selection of research proposals especially important regarding young scholars.

However, the attribution of grants is still an under-studied area. While studies on peer and panel review focus mostly on biases and fairness (see e.g. Lamont, 2009), we do not know how reviewers select the grant winners (van Arensbergen et al., 2014b). This is not only due to restricted knowledge on group decision making but importantly due to a lack of knowledge regarding what quality of research proposals means and how it can be identified. This is especially true for SSH disciplines (Hemlin, 1996).

In this talk we will present a method to develop an evaluation sheet for the ex-ante assessment of young humanities scholars' research proposals. The procedure follows a bottom-up approach and starts from the catalogue of quality criteria for humanities research (Hug, Ochsner, & Daniel, 2013). The criteria were adapted and expanded to the situation of ex-ante evaluation of young scholars' research proposals. This catalogue of criteria for research proposals was sent to the Swiss scholars in the humanities (theology/religious studies were excluded because of a similar survey of another project at the same time) who rated the criteria for their suitability for the evaluation of research proposals on a 6-point scale. Using the ratings, we identified criteria and aspects that reach a consensus among the scholars (50% of the scholars rated the aspect with at least a 5 and not more than 10% rate it negatively, i.e. with a 1, 2 or 3).

We randomly assigned the respondents to two groups, one group rated the criteria for the evaluation of proposals from PhD students, the second group rated the criteria for proposals from early career researchers at the postdoc level. The survey reached an overall response rate of 35%.

The first analysis of the results shows that there is high agreement among the three main groups of humanities disciplines: law, cultural history and language. From the 23 aspects specifying 9 criteria, only *societal relevance* and *cultural heritage* show important differences between law scholars and the others, *societal relevance* being of more importance to law scholars while *cultural heritage* being more pertinent for scholars from cultural history and language than for law scholars. However, none of the two aspects reaches consensus in any of the discipline groups. Overall, the following aspects did reach consensus among the scholars: *identifying gaps*, *innovative data* and *new findings* in existing paradigm (criterion originality); *timetable* and *resources* (feasibility); *research process*, *state of research*, *choice of method*, *argumentation* and *understandable* (rigour); *academic relevance* (relevance); *complexity* (complexity); and *variety of research* (variety). The following aspects did not reach consensus, overall: *independence* (independence); *new research topic*, *new approach*, *new paradigm* (originality); *societal relevance* (relevance); *cultural heritage* (cultural heritage); *CV*, *diploma*, *publications*, *recommendations* (applicant).

Some of these findings are in line with previous research on quality criteria for SSH research (e.g. a broad range of quality criteria is needed, societal relevance is not an indicator for quality, aspects of the criterion rigour reach the highest agreement, see e.g. Gogolin & Stumm, 2014; Hug et al., 2013; Lienhard et al., 2016). At the same time, some ratings show that the scholars were taking the evaluation situation into consideration: while new paradigm is considered an important criterion for research quality (Hug et al., 2013), in this survey it was considered not important because it cannot be expected from an early stage researcher. The same holds true for the aspects *new research topic* and *new approach* as well as *cultural heritage*, which usually can only be achieved with experience. Another striking result of the survey is the low rating of the aspects regarding the applicant. The scholars did not find that criteria about the person will inform about the quality of the proposal even though this is common practice.

In the presentation, we will further investigate differences between gender, groups of disciplines, status (tenured professor or not) and the level of the early career researcher (PhD or postdoc). We will then present an evaluation sheet based on the results of this survey of the scholarly community in the humanities as used in an actual evaluation situation at a Swiss university.

References

van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014a). Different views on scholarly talent: What are the talents we are looking for in science? *Research Evaluation*, *23*(4), 273–284. http://doi.org/10.1093/reseval/rvu015

van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014b). The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels. *Research Evaluation*, *23*(4), 298–311. http://doi.org/10.1093/reseval/rvu017

Gogolin, I., & Stumm, V. (2014). The EERQI peer review questionnaire. In: Gogolin, I., Astrom, F., & Hansen, A. (eds). *Assessing Quality in European Educational Research* (pp. 107-120). Springer VS: Wiesbaden.

Hemlin, S. (1996). Social studies of the humanities. A case study of research conditions and performance in ancient history and classical archaeology and English. *Research Evaluation*, *6*(1), 53–61.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*, *22*(5), 369–383. http://doi.org/10.1093/reseval/rvt008

Lamont, M. (2009). How professors think: Inside the curious world of academic judgment. Cambridge: Harvard University Press.

Lienhard, A., Tanquerel, T., Flückiger, A., Amschwand, F., Byland, K. & Herrmann, E. (2016). *Forschungsevaluation in der Rechtswissenschaft: Grundlagen und empirische Analyse in der Schweiz*. Stämpfli Verlag: Bern.

Are theories of peer review fit for the Humanities and Social Sciences?

Martin Reinhart

Peer review was and still is a prominent feature of modern science. As a central pillar for the allocation of scarce resources and the assurance of scientific quality, peer review has been remarkably polymorphic. From Henry Oldenbourg inventing the role of editor for the Philosophical Transactions in the 17th century to the plethora of current editorial review procedures, the forms and the diversity of peer review has evolved through time. From reviewing and deciding on manuscripts, project funding, and job applicants to elaborate forms of evaluation for funding programs, the performance of research units or complete (national) research systems, peer review currently shows noticeable adaptability (Weller 2001, Pontille & Torny 2014). How do theories of peer review explain such longevity and variability?

The literature on peer review has been notably silent on what adequate theories should entail, while at the same time producing a large quantity of empirical evidence on individual aspects of peer review (Hirschauer 2004). What can be called theories of peer review remains largely implicit in the literature or derives from folk theories and common knowledge of scientists reflecting on their individual experiences. The two most notable strands of theorizing derive from the premise that peer review is in essence either a decision mechanism or provides quality assurance. Very often, both are seen as compatible or even mutually supportive. In the last decade research has become more diverse and peer review has been conceptualized as an "inner public" (Hirschauer 2010), as a test (Pontille & Torny 2014), as a negotiation of epistemological differences (Mallard et al. 2009) by establishing fairness (Lamont 2009), as censorship (Biagioli 2002), or as an organized way to pacify scientific discourse (Reinhart 2012).

The mounting empirical evidence and the emerging theoretical concepts allow for taking stock and asking: What do current theories of peer review describe and explain in light of the available evidence? The first part of my talk will review empirical evidence and theoretical concepts in relation to each other by distinguishing between views of peer review that focus on outcome and those that focus on process (Abbott 2016). Theorizing peer review as an outcome provides a more intuitive understanding for the general scientific community and results in more numerous, and mostly quantitative, empirical studies. However, the resulting evidence, e.g. low reliability and indeterminable validity, gives grounds to criticize central assumptions of an outcome view of peer review. Theorizing peer review as a process, on the other hand, results in mostly qualitative case-studies. The resulting evidence seems very much in line with a processual view, but impedes generalization.

The second part of my talk will discuss this result on normative grounds and argue that a purely outcome based view of peer review will be detrimental to the Humanities and the

Social Sciences. While proposals and initiatives to reorganize peer review following an outcome model will seem sensible for certain research fields especially in the Natural Sciences, they may, however, have more problematic consequences for fields in the Humanities and the Social Sciences, where the peer review process remains an integral part of the capacity to produce knowledge. I will conclude by laying out the analytic and normative parameters of a processual view that aims at encompassing as much of the diversity of peer review as possible.

References

Abbott, A. (2016). Processual Sociology. University of Chicago Press.

Biagioli, M. (2002). From book censorship to academic peer review. Emergences: Journal for the Study of Media & Composite Cultures, 12(1), 11–45.

Hirschauer, S. (2004). Peer Review Verfahren auf dem Prüfstand: Zum Soziologiedefizit der Wissenschaftsevaluation. Zeitschrift für Soziologie, 33(1), 62–83.

Hirschauer, S. (2010). Editorial Judgments: A Praxeology of „Voting" in Peer Review. Social Studies of Science, 40(1), 71–103.

Lamont, M. (2009). How Professors Think: Inside the Curious World of Academic Judgment. Harvard University Press.

Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review. Science Technology & Human Values, 34(5), 573– 606.

Pontille, D., & Torny D. (2014). From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review. Human Studies, 38, 57-79.

Reinhart, M. (2012). Soziologie und Epistemologie des Peer Review. Baden-Baden: Nomos.

Weller, A. C. (2001). Editorial Peer Review: Its Strengths and Weaknesses. Medford: Information Today.

Databases and repositories for Social Sciences and Humanities research output: describing the present, discussing the future

Linda Sīle

During the last decades, national comprehensive databases have been set up in several countries; VABB-SHW in Flanders, CRIStin in Norway, PBN in Poland, and more. Their aim is to provide a complete national coverage of publications and other artefacts resulting from research in social sciences and humanities. Such databases promise, on the one hand, more trustworthy data sources for evaluative bibliometrics, but, on the other hand they are a rich source of insights into similarities and differences in publishing practices across countries.

Some of the databases, such as VABB-SHW, are well documented and the data contained have been studied extensively (e.g. Verleysen, Ghesquière, and Engels 2014; Engels, Ossenblok, and Spruyt 2012). In contrast, even basic information about other databases is relatively limited. For this reason, a study was launched in Autumn 2016 aimed at acquiring an overview of existing databases and repositories for social sciences and humanities research output in Europe and Israel. In this paper I report some of the findings of that study.

The study was conducted within the framework of the COST action 15137 "European Network for Research Evaluation in the Social Sciences and Humanities" (ENRESSH, http://enressh.eu/). The focus of the study was on national databases, but, given that such databases do not exist in all national contexts, attention was paid also to exemplars of other types of databases: databases and repositories within universities, national research agencies or other institutions. The data collection combined several methods. The main data collection instrument was a fact sheet covering various aspects of databases (e.g. data collection approach, included research output types). Additional information was acquired from scholarly literature, policy documents, and non-formal sources (websites, reports, e-mail interviews, etc.). The research participants were identified from ENRESSH and in own professional networks.

The study aimed to cover 41 countries in Europe, and Israel. Responses were received from 39 countries. So far, 24 of the databases reported can be considered national databases. It may seem that this is a surprisingly large number of national databases for social sciences and humanities research output. Indeed, the number of national databases depends on what we mean by 'national database for (social sciences and humanities) research output'. This is the topic that I will focus on in my presentation.

The study was launched, very much on purpose, without an explicit definition of a national database. However, implicitly the direction of the questions explored was oriented to those databases that are suitable for bibliometric research and research evaluation supported by bibliometric indicators.

The rationale behind such a strategy whereby no explicit definition of a database is mentioned is rooted in awareness of a possible risk that a definition, introduced without knowing the actual situation regarding research output metadata collection practices across countries, may lead to unnecessary exclusion of otherwise valuable data infrastructures. Indeed, the collected data do indicate a considerable diversity. Consequently, a suitable definition of a social science and humanities research output database can hardly be derived from them, a definition that would be inclusive yet meaningful. To give an example, does a national database for research output need to cover research output from all research-active institutions in the country or it is enough to cover output from only universities? Does it have to cover all possible research themes or is it sufficient to collect publication data only on themes regarded as central within the particular country? Questions such as these, on the one hand, point to unexplored cultural differences in bibliographic data collection practices, but, on the other hand, such discrepancies highlight the political nature of initiatives aimed at the integration of national data. Should this diversity be regarded as an asset or more as an obstacle to cross-country applications of evaluative bibliometrics?

One of the directions of the mentioned network ENRESSH is towards a design for a single European database for research output. Among the first tasks in such an initiative would be to find conceptually sound principles for integrating data from different countries into a single European database. However, the greatly varied ways of collecting and registering metadata on research output alert to the distinct possibility that the design of a single European database may produce new, or reinforce existing, inequalities. For instance, a decision could be made to include a particular set of publication types without a consideration that there may be a context where types that fall beyond the set are central. Consequently, conclusions drawn from such data may turn out to be inaccurate and biased towards certain academic contexts. This is the kind of inequality I have in mind. Yet, are categorial inequalities introduced through database design inevitable? With this question in mind I will discuss insights drawn from different databases for social sciences and humanities research output, searching for the principles needed for an optimally flexible design of a comprehensive database for European social sciences and humanities research output.

References

Engels, Tim C. E., Truyken L. B. Ossenblok, and Eric H. J. Spruyt. 2012. 'Changing Publication Patterns in the Social Sciences and Humanities, 2000–2009'. *Scientometrics* 93 (2): 373–90. doi:10.1007/s11192-012-0680-2.

Verleysen, Frederik T., Pol Ghesquière, and T. C. E. Engels. 2014. 'The Objectives, Design and Selection Process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW)'. *The Use and Abuse of Bibliometrics*, 115–125.

Towards integration of European research information

Hanna-Mari Puuska, Janne Pölönen, Tim Engels, Gunnar Sivertsen

Introduction

It remains a constant finding in bibliometric research that the most widely used commercial databases, Web of Science or Scopus, do not provide complete coverage of research output in any field, and that in SSH fields they suffer from severe lack of coverage of publications in books and languages other than English (e.g. Sivertsen 2016).

Science policy and research evaluation at all levels of the European Research Area need support from reliable, comparable, and comprehensive information on research activity, productivity and quality. To this end, an Expert Group on Assessment of University-Based Research recommended in a report to the European Commission that it should "Invest in developing a shared information infrastructure for relevant data to be collected, maintained, analysed, and disseminated across the European Union" (European Parliament 2010).

According to a report to the European Parliamentary Research Service (2014), 19 European Union Member States had developed or were developing a national research information system. The report recommends development of a European integrated research information system inter-connecting the existing national research information systems.

A database for the social sciences and humanities (SSH) outcomes is a crucial component of the European research information infrastructure. A European Scoping Project recommended, in a report to the European Science Foundation and the British, Dutch, French, and German research councils, that in order to achieve complete coverage of the SSH scholarly output, either negotiations to expand and/or create a new database with the suppliers of Web of Science, Scopus and Google scholar, or integration of data from national and institutional research documentation systems (Martin et. al. 2010).

The main difficulty of standardization and interoperability of data at the European level is the variety of national publication information systems and their data models. Many countries are facing a similar problem at national level when they compile information from research organizations using various local systems.

Finnish solution for integrating research information at national level

In Finland, an advanced decentralized solution to integrate institutional data at the national level, the VIRTA Publication Information Service, was launched in the spring 2016. VIRTA is a data warehouse, "a data hub" compiling bibliographic information annually of all scientific publications from 54 organizations using different local solutions for publication data collection, such as commercial Current Research Information Systems (CRIS), self-made publication registers, institutional publication repositories and e-forms.

VIRTA makes up-to-date metadata from research institutions available for other services and produce comprehensive and comparative information on publishing activity both nationally and institutionally. The publication metadata can be transferred to research funders, publication or data repositories, infrastructure services, or any other service used by researchers.

Applying Finnish concept to European level?

Adaptation of the VIRTA concept would establish a convenient and cost-efficient way for building a *European Publication Information Service* to which any European country or single organization would be able to provide its own publication data.

Figure 1. Publication information in European national and institutional systems usable for various purposes.

The planned European Research Information Service would provide a complete overview on European research publications including all types of scholarly publications and potentially other research outcomes in the future as well. Any country or institution would be able to join despite its own system architecture.

The information would also be usable in various processes such as researchers' CV's, funding applications, research evaluations, research administration, science policy planning, decision making, research and information retrieval. The data can be used in various digital services, for example, in importing publications to funding organizations reporting systems.

The European Research Information Service would provide an online infrastructure for research literature available for anyone through the public portal. The metadata of open access publications should always include links to full texts and thus provide a gateway to the actual research.

References

European Commission (2010). *Assessing Europe's University-Based Research: Expert Group on Assessment of University-Based Research*. Directorate-General for Research Science in Society 2008 Capacities, 1.4.1. https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf

European Parliamentary Research Service (2014). *Measuring scientific performance for improved policy making*. Science and Technology Options Assessment. European Parliamentary Research Service. April 2014. PE 527.383. http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET(2014)527383(SUM01)_EN.pdf

Martin, B., Tang, P., Morgan, M. Glänzel, W., Hornbostel, S., Lauer G., Lima, L., Oppenheim, C., van den Besselaar, P., Zic, M. (2010). *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*. A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF. 8 March 2010. https://globalhighered.files.wordpress.com/2010/07/esf_report_final_100309.pdf

Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities, *Scientometrics*, 107:2, 357–368.

Evaluation in Europe

SSH research evaluation in Europe: a classification

Michael Ochsner, Emanuel Kulczycki, and Aldis Gedutis

Evaluation practices differ widely across countries, and scholars have proposed different typologies of research evaluation systems (Coryn et al., 2007; Geuna & Martin, 2001;2003; Hicks, 2010; 2012; von Tunzelmann & Mbula, 2003). However, none of the observations focused on SSH research in detail. Furthermore, only a small number of countries are included in the typologies, mostly those for which information on evaluation practices is available and widely discussed, such as the United Kingdom (RAE/REF), the "Norwegian system", Belgium (Flanders), or the Netherlands. Therefore, not much is known how SSH research is evaluated in European countries, nor about similarities and differences between the evaluation systems in the different countries.

In this presentation, we suggest a classification that is useful for most European countries, for some of which there is no information available yet to the international community, and that focuses on evaluation of SSH research. To develop such a classification, we chose an iterative design. The goal is to find the common elements in the different typologies, terminologies and dimensions along which we can classify and compare the countries with regards to their evaluation systems.

We applied a Delphi-like approach to be able to reflect the potential peculiarities of SSH research evaluation and of countries of very different backgrounds (for the use of the Delphi method to create a typology of evaluation systems, see Coryn et al., 2007; for a Delphi-method in the context of SSH research evaluation, see Hug, Ochsner & Daniel, 2014). In a first step, we developed an initial typology identifying several characterising dimensions based on existing typologies of research evaluation systems. In a second step, we administered a questionnaire based on this initial typology to the specialists of the COST-Action "European Network for Research in the Social Sciences (ENRESSH)". This questionnaire was developed by members of the Steering Committee and selected specialists from the Management Committee of the Action. The purpose was not yet to classify the countries but rather to optimise the typology, assure that it covers a wide range of countries and to test the consistency of the classification among the respondents from the same country. The results and the feedback was used in a third step to adapt the typology within the Work Group "Conceptual frameworks of SSH evaluation" of the same COST-Action. In a fourth step, we administered the adapted questionnaire to the specialists of the enlarged COST-Action (the COST Action grew from 22 to 35 countries in the meantime). Using this information, we classified the different countries. At the time of writing this abstract, the questionnaire is in the field and results are expected in March/April.

The first questionnaire consisted of nine dimensions: level of the evaluation protocol; differentiation; who is evaluating; object of evaluation; funding; method; timeline; transparency; and costs. Some of the dimensions are further specified by aspects, e.g. the dimension differentiation is divided into a) differentiation of SSH vs. STEM evaluation

procedures and b) differentiation of academic universities of universities of applied sciences. The results revealed that the dimensions of the existing typologies do not suffice to adequately describe SSH evaluation systems: First, the added dimensions (e.g., differentiation, transparency and cost) discriminate between evaluation systems (there is variance between countries); second, the comment fields were used and new dimensions suggested. Of course, this is also due to the more heterogeneous selection of countries included in this study than in the existing typologies. Furthermore, the responses to the questionnaire also showed clearly that the term "evaluation system" is not clear and answers to the questions included a diverse range of evaluation situations from ex-post institutional evaluation (the intended meaning stated in the introduction of the questionnaire) to national systems of career promotion to ex-ante project evaluations. This showed clearly that national evaluation systems are quite heterogeneous and it is worthwhile to create a classification that takes such national differences into account.

Besides these general results regarding the further development of the questionnaire, the responses to the survey reveal that in many countries an accountability-based evaluation of SSH is in place: In the majority of the countries, evaluation is organised at the national level (19 out of 25 countries). In 13 of the countries, evaluation is also related to funding. If we add those who agree to the statement "officially, evaluation is used to provide feedback (formative evaluation), but funders or universities base their funding decisions on evaluation outcome", the number rises to 15 countries.

With regard to SSH evaluation, only 18 out of 43 respondents from 14 out of 24 countries affirm that evaluation procedures for SSH research are adapted to the SSH. However, only one scholar (and thus one country) affirms that there is no use of citation data for evaluating SSH research. This is surprising because such an answer would have been expected to be more often linked to "evaluation of SSH disciplines is SSH specific".

While these are only first results and there is some disagreement between respondents from the same country for some dimensions, the results suggest that evaluation of SSH research is not always SSH-specific even though the evaluation specialists strongly encourage discipline-specific procedures. And if evaluation is SSH-specific, is seems to be so because standard practices fail to provide meaningful results rather than because it was carefully designed to reflect SSH research practises and goals.

To further investigate these results and to find a classification that adequately reflect the diversity of European research evaluation practices, the first questionnaire was reworked into a second wave questionnaire. It added aspects to dimensions and also reflects the three different types of evaluation: institutional ex-post evaluation, career promotion and ex-ante project evaluation.

The questionnaire will be fielded between March and May, and the presentation at the RESSH conference will thus focus on the results from this second Delphi wave. We will sketch a picture of European SSH research evaluation and highlight differences and commonalities between countries and groups of countries.

References

Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). "Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison", in *American Journal of Evaluation*, 28(4), 437–457, http://doi.org/10.1177/1098214007308290.

Geuna, A., & Martin, B. R., (2001). "University Research Evaluation and Funding: An International Comparison", in *SPRU Electronic Working Paper Series* (Vol. 71).

Geuna, A., & Martin, B. R., (2003). "University Research Evaluation and Funding: An International Comparison", in *Minerva*, 41(4), 277–304. http://doi.org/10.1023/B:MINE.0000005155.70870.bd.

Hicks, D. (2010). "Overview of models of performance-based research funding systems", in *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD Publishing, 23–52. http://doi.org/10.1787/9789264094611-4-en.

Hicks, D. (2012). "Performance-based university research funding systems", *Research Policy*, 41(2), 251–261. http://doi.org/10.1016/j.respol.2011.09.007.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2014). "*A framework to explore and develop criteria for assessing research quality in the humanities", in International Journal of Education Law and Policy*, 10(1), 55–68.

von Tunzelmann, N., & Mbula, E. K. (2003). *Changes in research assessment practices in other countries since 1999: final report*.

The contradictions of the European (Open) Science policies with regard to the evaluation of research and publications in the social sciences and the humanities

Marc Vanholsbeeck

The European Research Area (ERA) acts as an important provider of policy prescriptions in the field of research and innovation, now fully integrating the social sciences and humanities (SSH) (Birnbau et al. 2017) and inspiring policy making at national and institutional levels with regard to a diversity of thematic priorities (Commission 2012). European prescriptions are of a political as well as a managerial nature, setting objectives to be achieved and prescribing managing tools to monitor the progresses made towards the completion of these objectives. Since the EU has until now never used the legislative power on which it could rely, European prescriptions are mostly articulated through soft laws -, policy documents and statements as well as, more indirectly, through the management tools of the ERA and the rules that govern the research and innovation funding programs.

On the basis of a qualitative analysis of these political and managerial prescriptions and a first-hand quasi-participant observation of the European research policy making, we will show how European prescriptions have contradictory implications for the evaluation of the diversity of research outputs that characterize SSH research and, more particularly, with regard to the types of SSH publications which have a particular potential to impact society.

On the one hand, ERA has been strongly influenced in its prescriptive activities since the 1990s by programmatic ideas such as the "mode 2 of knowledge production" (Nowotny et al. 2003) and is increasingly supporting the "exoterization" of research beyond the traditional "esoteric" circles of the peers. Hence ERA prescribes that research activities will target – and even include in a "co-creation" perspective - a broader audience of innovators and citizens, through the use of concepts such as "societal challenges" – which constitute the best funded of the three pillars of the current Horizon 2020 framework programme -, RRI ("Responsible Research and Innovation"), "Science With and For Society" or "Citizen Science". For SSH researchers, an important way to be in line with this prescribed tendency towards exoterization is to deliver research outputs and types of publication that may reach an audience beyond Academia, such as: publications in vernacular languages, monographs, policy briefs, policy reports, results from locally anchored (action) research, textbooks, translations, exhibition catalogues, popularization works or book-length scholarly essays (Hicks 2005).

On the other hand, European prescriptions concurrently support tendencies towards a NPM-driven "managerialization" of research that uses publications as regular indicators of performance, and bibliometric international databases (Web of Science and Scopus) as important management tools (Vinkour 2014). Together with the more general trends towards massification and internationalization of research, this last tendency has contributed to foster an evaluation context in which research quality is increasingly assessed in light of the number of articles in International Top Journals, often published by Majors (Larivière

et al. 2016), targeted towards academics – thus mostly inaccessible to exoteric audiences -, and in which a large share of the aforementioned SSH research outputs is not adequately taken into account (Archambault et al. 2006).

Recent EU policies promoted under the banner of "Open Science" are to be deemed as ambivalent in regards with this contradiction between exoterization and managerialization of research in the fields of SSH, presenting opportunities for its resolution, but also some risks of further extending it. Open Science is indeed framed in European policy documents and statements – such as in the discourses of the current Commissioner for research, science and innovation Carlos Moedas - as the epitome of the participative approach to research, with explicit support to the development of alternative metrics and the assessment of research impact beyond Academia. Though the stress is concurrently put at EU level on achieving a total flipping to an Open Access (OA) dissemination model before 2020 (Council 2016) – with a focus on articles of peer reviewed journals and embargo periods as short as possible –, while giving an implicit support for the negotiation of "big deals 2.0" with Major publishers (Chartron 2016). In our experience, this particular approach towards OA, while pragmatic and promising in the short term, is at risk of reinforcing the imbalances inherent to the use of international bibliometric databases in the evaluation of SSH research. It is not in favour either of a better archiving, visibility, accessibility and assessability of those aforementioned SSH research outputs that, while not being published in International Top Journals, are best placed to get an impact on society.

Hence we believe that a more systematic, inclusive and complementary use of OA repositories, through which the whole scope of SSH research production could be made public and accountable – i.a. through a diversity of metrics and altmetrics -, together with the recognition of a broader definition of what the very notion of publication means in the SSH, may contribute to overcome the aforementioned contradiction in European science and Open Science policy making. In this perspective archiving all types of publication in OA repositories will allow SSH researchers to be better rewarded, in evaluation situations, for the whole scope of the contributions they make, as scholarly citizens, for the benefit of our society.

References

Archambault, E., Vignola-Gagne, E., Côté, G., Larivière, V., Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. Scientometrics, 68(3), 329-342.

Birnbaum, B.I., Keraudren, P., Strom T., & Vavikis, T. (2017). *Integration of social sciences and humanities in Horizon 2020: participants, budget and disciplines, 2nd monitoring report on SSH-flagged projects funded in 2015 under the Societal Challenges and Industrial Leadership priorities.* Brussels: DG RTD.

Chartron, G. (2016). Stratégie, politique et reformulation de l'open access. Revue française des sciences de l'information et de la communication [En ligne], (8). Retrieved January 10, 2017 from http://rfsic.revues.org/1836

Council (2016). *The transition towards an Open Science system - Council conclusions adopted on 27/05/2016.* Retrieved February 28, 2017 from: http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

European Commission (2012). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Reinforced European Research Area Partnership for Excellence and Growth (17 July 2012). Retrieved December 23, 2013 from http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf

Hicks, D. (2005). The Four Literatures of Social Science. In H. Moed, W. Glänzel, & U. Schmoch (Ed.), *Handbook of Quantitative Science and Technology Research* (pp. 473-496). Dordrecht: Springer.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. PLoS ONE, 10(6). e0127502. doi:10.1371/journal.pone.0127502.

Nowotny, H., Scott, P., & Gibbons, M. (2003). Introduction: Mode 2 Revisited: The New Production of Knowledge. Minerva, 41(3), 179-194.

Vinkour, A. (2014). La normalisation de l'université. In B. Frydman, & A. Van Waeyenberge (Ed.), Gouverner *par les standards et les indicateurs: De Hume aux rankings* (pp. 235-262). Brussels: Bruylant.

Local organizing committee

Tim Engels

Kristina Røtterud Granaas

Raf Guns

Linda Sīle

Karen Van Camp