VALORIZING SSH RESEARCH: TOWARDS A NEW APPROACH TO EVALUATE SSH RESEARCH' VALUE FOR SOCIETY

IOANA GALLERON, MICHAEL OCHSNER, JACK SPAAPEN AND GEOFFREY WILLIAMS

In the last decades, we have witnessed a shift towards accountability and new public management practices in the management of universities in most countries. (Hamann, 2016; Hammarfelt and De Rijcke, 2015; Kekäle, 2002; Mali et al., 2016). Due to the pressure to be efficient and accountable, universities have implemented comprehensive evaluation procedures for research performance and research impact (Geuna & Martin, 2003). In addition, the availability of quantitative data and the preference of managers to use numbers to compare performance, has led to evaluation systems that are mostly based on science indicators, either drawing on data from Thomson Reuters' Web of Science or measuring direct, and sometimes indirect, economic effects. At first, this mainly affected the Science, Technology, Engineering and Mathematics (STEM) disciplines, but the budget constraints following the 2008 global financial crisis and its aftermath, as well as the continuing demand for accountability, led to wider implementation of quantitative research assessments, including the Social Sciences and Humanities (SSH) (Burrows, 2012; Guillory, 2005). However, while such campaigns led to considerable results in some cases, improving the overall performance of certain research systems, they were all confronted with numerous difficulties when it came to the evaluation of the SSH. A much broader resistance has developed against the quantitative approach of research quality and impact, also in the STEM-disciplines, as witnessed by the San Francisco Declaration on Research Assessment (DORA).

Research evaluation has always been perceived as a difficult area for the SSH for various reasons, amongst which being the wide variety of disciplines, approaches and practices brought together under the umbrella term of SSH. Problems mostly arise from the fact that the most common procedures have been fine-tuned to hard sciences and their production and communication practices, and as such they are ill-adapted to the research practices, to the national variations and to the dissemination traditions in the SSH disciplines. No wonder that a certain reluctance grew in the SSH fields, all the more so since SSH scholars believe strongly in the value of their disciplines for the advancement of knowledge, and in the contributions they can make to education, culture, the political system, work related issues, etc. As a consequence, the development of assessment procedures that are able to adequately review the work of SSH researchers have become necessity.

This paper endeavours to present the rationale for a valorising evaluation of SSH research. It will start with a presentation of the results of an initial survey about SSH research evaluation in Europe, conducted within the COST Action 15137 ENRESSH (the European Network for Research Evaluation in the Social Sciences and Humanities). The survey allowed us to capture perceptions about the main principles informing SSH research evaluation, to advance towards a typology of different evaluation models, and to better identify the problems linked to these models. In the second part, looking at good practices and based on these preliminary results, we will present an approach that combines performance and impact in a way that can represent a solution for SSH evaluation, and possibly beyond. In the third part, we will move to an analysis of the difficulties of valorising SSH through evaluation. These have been discussed to a certain extent in the recent literature (Spaapen & van Drooge, 2011; Molas-Gallart & Tang, 2011; de Jong et al., 2014), especially after different countries adopted evaluation campaigns introducing the criteria of "societal impact". However, while these discussions concerned questions such as attribution of impact to a specific research endeavour, or technical difficulties, such as how to document impact (Penfield et al., 2014), this paper will look more at the challenges of a fine understanding of the exact place of SSH in science and society, as a prerequisite for their evaluation.

I. TOWARDS A TYPOLOGY OF SSH RESEARCH EVALUATION SYSTEMS IN EUROPE

It is well known that evaluation practices differ widely across countries, and, over time, scholars have proposed different typologies of research evaluation systems (Coryn et al., 2007; Hicks, 2010; 2012; Martin & Geuna, 2001; 2003; von Tunzelmann & Mbula, 2003). However, none of the observations focused on SSH research evaluation in detail. Moreover, typologies have to date focused on a small number of countries, mostly those for which information on evaluation practices is available and widely discussed, such as the United Kingdom (RAE/REF), the "Norwegian system" (based on CRIStin), the evaluation in Belgium/Flanders (based on VABB_SHW), or the protocol for evaluation used in the Netherlands (SEP 2015-2021).

Given this lack of broad knowledge on SSH evaluation procedures, one of the first endeavours of ENRESSH was to observe and compare how research in the SSH is evaluated in different countries. The focus of the project is on European countries, even if, as the ENRESSH network expands, we start to be able to gather insights as to how the SSH are evaluated more widely.

To create a typology along which the countries can be classified according to their evaluation systems, a Delphi-like approach was adopted (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). The procedure consists of five steps. In a first step, a provisional typology was developed by the members of the Steering Committee and selected specialists from the Management Committee of the Action. In a second step, a survey based on this typology was administered to the specialists of the COST Action. The purpose was not at this stage to classify the countries, but rather to optimise the typology and to test the consistency of the classification among the respondents from the same country. The results and the feedback are being used in a third step to adapt the typology and to build an adapted questionnaire, that will be administered again to the specialists in a fourth step. Finally, the results as well as supplementary documents will be used to classify the evaluation systems of the countries. In the following, we are reporting results from the first two steps.

For the development of the initial typology, we started with a literature review, allowing us to identify several characterising dimensions on the basis of existing typologies. To these, we added some aspects we found were missing and/or specific to the SSH. As a result of this process, we designed a first typology consisting of the following dimensions: level of the evaluation protocol; differentiation; who is evaluating; funding; method; timeline; transparency; and costs. They are described in more detail as follows:

1. Evaluation is organised at different *levels* (von Tunzelmann & Mbula, 2003). Some countries have a national evaluation system, while in other countries, evaluations are organised at the regional level or subject to each university's autonomy. We differentiated between the level of organisation of the evaluation system on the one hand and the level on which data for evaluations are collected (existence of national, regional or institutional databases).

2. Research practices and communications in the SSH differ in a number of ways from research in the STEM disciplines. For example, commonly used evaluation practices, e.g. bibliometrics based on Web of Science data, are not readily applicable in the SSH (see, e.g., Hug, Ochsner & Daniel, 2013; Hicks, 2004; Nederhof, 2006; Ochsner, Hug & Daniel, 2012) and similar issues also arise with applied research (Furlong & Oancea, 2005). As a result, an additional dimension that is not yet present in the existing typologies[1] must be added: *differentiation*. It includes two aspects: a) whether there are specific methods or procedures to evaluate SSH research and b) whether there are different evaluation procedures for applied and for basic research.

3. Different *bodies* can be responsible for conducting or supervising the evaluation (Geuna & Martin, 2003; Hicks, 2010; 2012; von Tunzelmann & Mbula, 2003). Sometimes the differentiation between the level on which the evaluations are organised and the body responsible for evaluation is not very clear (von Tunzelmann & Mbula, 2003).

4. Evaluation can be linked to *funding* or serve formative reasons (Coryn, Hattie, Scriven & Hartmann, 2007; Geuna & Martin, 2001; 2003; von Tunzelmann & Mbula, 2003). Obviously, there is also the possibility that the evaluation outcome is not officially linked to funding but is nevertheless used for funding purposes by other institutions, or inside the evaluated institution.

5. Different *methods* can underlie the evaluation procedure (Coryn et al., 2007; Geuna & Martin, 2003; Hicks, 2010; 2012; von Tunzelmann & Mbula, 2003). This dimension has the following aspects: a) the principal method; b) whether and what kind of data is used; and c) criteria that are used if peers are involved.

6. Evaluations involve a time dimension. Two aspects are linked to time: a) evaluations can be repeated, thus the *time* of an evaluation cycle (in other words whether it is consistent and systematic) is a first aspect (Coryn et al., 2007; Hicks, 2010; 2012); and b) evaluations look back at a certain time window, which constitutes a second aspect (Hicks, 2010; 2012).

7. *Transparency* is an important dimension regarding dissemination and the use of the evaluation (Dahler-Larsen, 2012; Hammarfelt, Nelhans, Eklund & Astrom, 2016; Hicks, 2010). This is closely linked to the method applied and whether there is a link to funding. As results of evaluations can be seen as indicators of quality themselves, transparency and a reflective dissemination are crucial. Evaluations engage therefore an ethical responsibility (see Hicks et al., 2015; Klein, 2008); it is also a requirement for the construction of indicators in general (see the OECD Handbook of Composite Indicators, Nardo, Saisana, Saltelli & Tarantola, 2005), as well as in program evaluation (Morris, 2015). However, while Hicks mentions that "Most systems emphasise transparency of methods and data" (Hicks, 2010, p. 39), she does not use transparency for the typology. In our case, we use three aspects of transparency: a) the methods for calculating the final scores, when these are an outcome of the evaluation; b) the methods for linking scores to funding, if funding is linked to evaluation; and c) the publication of the results.

8. Evaluations come with a *cost* in both time and money. Hicks emphasises the need to include the cost of evaluation in a typology, but states that "cost is rarely discussed" (Hicks, 2010, p. 34). She observes also that assessing "costs and benefits [...] is impossible" (Hicks, 2012). Geuna and Martin (2003) also raise the question of the cost/benefit ratio of performance-based research assessments. However, they did not investigate whether this was a topic in the countries. Rather, they argue that such systems, in general, will not have a positive cost/benefit ratio in the long run as the procedures become more and more complex and the returns on investment diminish as more countries apply the same procedures. We included two aspects regarding the costs in our typology: a) whether (estimated) costs are made public and b) whether there are efforts to estimate cost/benefit ratios.

---

1   Hicks (2010; 2012) mentions that field-specific approaches are necessary and states that "all systems are sensitive to differences in the patterns of fields' output" (Hicks, 2010, p. 37). In particular, the SSH are to be treated differently. However, she does not classify or specify how the systems account for differences.

A questionnaire based on these dimensions and aspects was then devised and administered, in March/April 2016, to the sixty members of the Management Committee (MC) of the COST-Action. The latter are all experienced in topics related to SSH research evaluation and represent their countries in the Action. The purpose of the survey was to get a first impression as to how the dimensions are used by the representatives of the countries to describe the evaluation system in their country.

Despite the time constraints – the fieldwork lasted for less than one month – , 43 persons from 25 countries filled in the questionnaire, of which 36 respondents from 22 countries answered all the questions. Countries were represented by one to five respondents; ten countries were represented by more than one respondent at least for a part of the questionnaire. The 25 countries in the study were: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, France, Germany, Iceland, Ireland, Italy, Lithuania, Macedonia, Malta, Republic of Moldova, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Switzerland, and United Kingdom.

The results confirm that the dimensions of the existing typologies do not suffice to adequately describe SSH evaluation systems. First, there is variance between countries in the dimensions and aspects we added, e.g., differentiation, transparency and cost. Second, the intensive use of the comment fields showed that even more dimensions or aspects should be taken into account in order to adequately reflect the different evaluation systems. Obviously, this is also due to the more heterogeneous selection of countries included in this study as opposed to the selections on which other existing typologies are based.

While there was agreement between representatives of the same countries regarding the *methods* applied, *who* is responsible for the evaluation and whether results are used for *funding* decisions, there was much disagreement regarding the other dimensions. This disagreement might be due to a number of reasons. For example, the comments showed that while the survey had set out to tackle national evaluation systems regarding ex-post research evaluation, the respondents had all kinds of evaluations in mind, from ex-ante evaluations of research proposals to appointments to professorships and ex-post evaluations. Some also mentioned that they differentiate between evaluation and assessment, in terms of defining evaluation as formative and assessment as linked to funding. While this might be due to an inadequate definition of the terms in the survey, we rather interpret this as reflecting the national differences in the organisation of evaluation. For instance, in some countries appointments to professorships are organised nationally, and thus were included in the responses to the survey by the representatives of these countries, while in others, appointments are organised at the institutional level, and thus not subject to this survey for the representatives of those countries. Bearing this in mind, rather than being restrictive in our definitions, we are gradually adapting the dimensions and aspects of the typology, so as to take into account these national differences.

Besides these insights into the national differences of evaluation systems, the questionnaire confirms the existence of an accountability-based evaluation applied to the SSH in many countries. More often than not, evaluations are national: in 19 out of the 25 countries covered by the survey, respondents report a form of national evaluation, whether it be institutional or individual, and this proportion remains high even if we exclude the three cases with a strong disagreement between respondents from the same country (Belgium, Croatia and Spain). 20 respondents from 13 countries affirm evaluation is related to funding; these figures become 25 and 15 when one adds those respondents considering that

"officially, evaluation is used to provide feedback (formative evaluation), but funders or universities base their funding decisions on evaluation outcome". This is to be corroborated with the answers from six countries according to which evaluation is "solely for feedback", as well as with the fact that the degree of agreement between respondents from the same country is higher than for other questions.

Interestingly enough, the situation is ambiguous regarding the differentiation dimension. Only 18 out of 43 respondents from 14 out of 24 countries affirm the exercises to be adapted to the SSH; four respondents perceive an evaluation to be SSH-specific if no citation data is used for some of the SSH, but affirm that, otherwise, the same procedures are applied. This is to be corroborated with the fact that only one scholar from one country affirmed that there is no use of citation data for evaluating SSH research, an answer one would have expected more often linked to the answer "evaluation of SSH disciplines is SSH specific". In short, one gets the impression that the evaluation of the SSH is not always SSH-specific, even though research on evaluation strongly encourages discipline-specific procedures. Furthermore, if the evaluation is SSH-specific, it is so because of the failures of existent procedures (e.g. bibliometrics cannot be applied to the SSH), rather than because it was carefully designed to reflect SSH research practices and goals.

2. FROM ACCOUNTABILITY TO VALORISATION

Is it possible to overcome the shortcomings of the existing evaluation protocols applied to the SSH, as observable through the responses quoted above, and to propose an intellectual frame, as well as methods and techniques truly adapted to these disciplines? The way forward seems to be a shift from the main principles, frameworks and practices of current research evaluation, as schematically described above, towards an approach looking for a combination of performance and valorisation of research in these disciplines. This does not mean pleading for a one-size-fits-all approach, nor abandoning the criteria of scientific quality, and even less forgetting about the accountability of sciences to the society. The idea is to reorient the evaluation exercises in a way that would be both more acceptable to the SSH scholars themselves and also able to provide a much needed evidence for informed decision making. This has the advantage of allowing specialists in the research evaluation to concentrate upon the numerous questions that arise, instead of trying to adapt "traditional" evaluation methods, often metrics based, to the specificities of SSH research, a somewhat procrustean endeavour.

A valorisation model for evaluation starts from the assumption that SSH research produces value, both for academia and for society, and that a large part of this value is not measurable in quantitative terms, nor assessable in other tangible terms. SSH research often regards new perspectives and insights that may influence the organisation and structure of processes and sectors in society. Whether regarding the "hard" sciences or, in more recent analysis, regarding SSH research, it has been repeatedly demonstrated that "impact" does not repose on a linear model, and that major innovations, be they technological, economic or societal, are multifactorial and cannot be related with certitude to a specific research project, publication or team (Greenlagh et al, 2016; Bornmann, 2012; Bornmann, 2013).

Nevertheless, the assumption that the SSH produces value is far from

being a matter of pure belief. Recent developments have shown, one more time, that there is a continuum between "hard" and "soft" sciences (Desmond Hellman, 2016), and that underfunding or undervaluing of the latter may hinder important developments in the first when they are much needed (Bod, 2013). Also, what education brings to society cannot be easily measured, but there is much evidence that education in all kind of subjects (even in "obscure" disciplines and fields of research, such as rare and ancient languages, for instance), and not mere training in immediately employable, job market needed tasks, is the basis of an articulated democracy (see Nussbaum, 2010).

Consequently, a valorisation model should concentrate less upon the "value for money" dimension and more upon finding the ways to stimulate the production and the dissemination of SSH knowledge. Considering that the uptake of research advancements is uncertain and not programmable, a model should pay less attention to "impact" understood as "modification in B due to A", and more to the collaborative dimension in these disciplines, even if this means inviting many of them, traditionally characterized by a solitary hermeneutical approach, to a considerable epistemological shift. If the goal is to get the most from SSH research, a valorisation model should set criteria and standards by stimulating strongly connected SSH, both to academia and to society. In other words, evaluating to valorise involves understanding and rewarding high quality, interdisciplinary and societally connected research, rather than concentrating on either academic or societal impactful research. This has the advantage of evaluating scholars, teams or institutions on the basis of what they actually do (or not), including the pro-active and innovative ways they develop to engage with the scientific community and society, rather than on the basis of what the scientific community and/ or the society does (or not) with their research. While societal and academic relevance should always be pointed out for any research undertaken, it is important to understand that the actual impact cannot be demonstrated in an unrealistically short time-frame and using questionable evidence. Most impacts that really make a difference may take 10 to 15 years. This understanding may also help to prevent perverse effects such as focusing only on research that comes with low risk and with short-term attention in academia and society. At the same time, it may also result in a certain slowing of the race to publication and citation, and may alleviate the burden of collecting "proofs of impact" which weighs heavily on researchers from certain countries, to the detriment of the time they can actually dedicate to research and teaching itself. Having said that, it does make sense to look at short and medium term effects in the context of the larger innovation process, for example via contributions of researchers to that innovation process.

Fortunately, some large-scale experimentations of a valorising evaluation model have been conducted and are leading to an assessment of (SSH) research in more understanding ways, with both the scientific quality and the societal relevance assessed. In the literature, many examples of new approaches for evaluation of societal impact can be seen (see special issue of *Research Evaluation*, September 2011; RAND 2013; several HEFCE reports; Lyall 2013). Practical examples can be found in the Netherlands, where a comprehensive framework specifically for humanities research has been presented (https://www.qrih.nl/nl/)[2]. More methods around impact pathways are currently developed in the second working group of ENRESSH.

In short, novel approaches to evaluation need to include the following:
- Knowledge about SSH research production;
- Some form of socially distributed responsibility, stakeholder involvement;
- Focus on the context of application of knowledge, next to scientific excellence;
- Be subject to multiple accountabilities (collegial/professional vs. managerial).

3. DIFFICULTIES OF VALORISING EVALUATION OF SSH RESEARCH

INTRODUCING A VALORISING MODEL DOES NOT GO WITHOUT ITS OWN DIFFICULTIES.

While standards of quality are controversial in all disciplines, recent research shows that perceptions and conceptualisation of excellence are even more complex and fuzzy in the SSH (see, e.g., Furlong & Oancea, 2005; Hemlin, 1996; Ochsner, Hug & Daniel, 2014; Williams & Galleron, 2016). Also, while peer-review is generally universally acclaimed and accepted within this area, in many journals or publishing houses, as well as at other levels and institutions where evaluation is practiced by peers, procedures are far from being transparent and robust, and often have not been closely monitored or assessed against principles such as thoroughness and fairness (Hemlin, 2009). Moreover, the relationship between science and society is changing and evaluation mechanisms are bound to reflect that to a certain extent. Boundaries between the two spheres become blurred and stakeholders become more involved in collaborations with researchers. This means that in some cases their interests and goals have to be included in review systems, also in the peer review, or expert review as it becomes then (Hemlin, 2006).

Another difficulty is that "societal impact" as a concept is difficult to define since it depends heavily on the context. If we limit ourselves to SSH research, it is clear that a researcher doing work, for example, in the area of religious studies working on the integration of Muslims in Western societies will be working in a rather different context than a researcher who is working in history of technology aiming at a new curriculum for high schools. The former most likely has to collaborate with people from religious and other communities and with policy and law makers, while the latter might work in the context of secondary education. Debates will differ, and so will the needs of these different stakeholders. This affects the kind of products needed by stakeholders. In the case of religious studies, knowledge exchange and policy proposals might be a prime goal, in the case of history of technology, a course or a book might be the product. Regarding evaluation, this means that it is difficult to come up with societal impact measurements that are adequate for all or most fields in SSH. Moreover, while the word 'impact' has a linear connotation (with a sender and a receiver) these two examples confirm what has been suggested in the second section, that results can only be achieved in interaction with stakeholders, which has to be stimulated more than the actual impact being measured. A new course in history of technology cannot be developed without the school community that

---

is supposed to work with it. And the knowledge developed by scholars who study religion can only help the debate about the migration crisis through a debate with other parties involved.

In discussing societal impact, much use is currently made of the quadruple helix model, in which government, industry, academia and civil participants work together "seamlessly". Such a biological metaphor raises several questions when put in the context of research evaluation. The biological double helix discovered by Watson and Crick only works because there is interaction between the strands, which are not just running in parallel (and keep working because there is RNA that detects and repairs flaws); the same applies to the quadruple helix model where it is vital to clarify what these strands are and what they seek to achieve. The first problem that arises is in deciding who stakeholders actually are, and then trying to identify their motivations, perceptions and goals as each of the current strands actually covers very different entities. The obvious stakeholders are the scholars themselves, but their objections against STEM geared evaluation approaches have rarely been taken seriously up until now. A second group of stakeholders is that of the policy makers and funders, which might be as heterogeneous as the SSH themselves. A third group is society, even more diverse: public organisations, NGO's, small and big industry, and the public at large, who mostly values the SSH for the cultural knowledge and wisdom that underlie stable democracies where freedom of thought is cherished. An analysis of all these strands is necessary to find out where and how they are connecting, and where they are not. Understanding the strands means raising awareness across the board so that common ground can be found. ENRESSH is already working towards this, with the aim to create a dialogue between different policy makers as well as opening up debate as to other aspects of the helix.

CONCLUSIONS

Evaluating to valorise is particularly important for the SSH disciplines. However, much research is still needed in order to proceed towards such a model. An in-depth understanding of SSH knowledge production processes and strategies is needed as a basis for developing evaluation procedures that adequately reflect the research practices, goals and aims of the SSH scholars. In parallel, the engagement of SSH researchers with societal challenges has to be attentively studied, so as to have a more comprehensive view of the ways in which interaction takes place in non-academic partnerships and environments of SSH research. Lastly, robust data about SSH production has to be gathered, and this means in many cases creating from scratch research information systems dedicated to SSH research outcomes.

ENRESSH seeks to accelerate progress on all these topics, through coordinating research projects going on in several European countries. While primarily aimed at reorienting the evaluation of SSH research, its results may prove useful for the entire of academia, as voices are numerous in the STEM sciences pointing out that this area is also diverse, that many disciplines are ill-served by a "one-size-fits-all" approach, and that evaluation driven by a narrow set of scientific excellence criteria and/ or demands of "usefulness" does not do justice to the wealth of contributions research is bringing to the advancement of knowledge and to the society.

REFERENCES

Bod, R. (2013). A New History of the Humanities. The Search for Principles and Patterns from Antiquity to the Present, Oxford, Oxford University Press.

Bornmann, L., (2012). "Measuring the societal impact of research", EMBO reports, vol. 13, no. 8.

Bornmann, L. (2013). "What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey", Journal of the American Society for information science and technology, 64(2), 217–233.

Burrows, R. (2012). "Living with the hÐindex? Metric assemblages in the contemporary academy", The Sociological Review, 60(2), 355–372, http://doi.org/10.1111/j.1467-954X.2012.02077.

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.

Dahler-Larsen, P., (2012). "Evaluation as a Situational or a Universal good? Why Evaluability Assessment for Evaluation systems is a Good Idea, What it Might Look Like in Practice, and Why it is not Fashionable", Scandinavian Journal of Public Administration, 16/3: 29–46.

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", in Research Evaluation, 23(2), 89–102.

De Jong, S., van Arensbergen, P., Daemen, F., van der Meulen, B., van den Besselaar, P., (2011). "Evaluation of research in context: an approach and two cases", in Research Evaluation, 20(1), 61-72.

Desmond-Hellman, Sue (2016), "The Latest Weapon in The Fight Against Disease Might Surprise You", published on October 26, 2016 on Linkedin.

Furlong, J., Oancea, A., (2005). "Assessing quality in applied and practice-based educational research: A framework for discussion". Oxford University Department of Educational Studies. Retrieved from http://www.esrc.ac.uk/ESRCInfoCentre/Images/assessing_quality_shortreport_tcm6-8232.pdf

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.

Guillory, J. (2005). "Valuing the humanities, evaluating scholarship" Profession (MLA), 11, 28–38.

Hamann, J. (2016). "The visible hand of research performance assessment", Higher Education, 1–19. http://doi.org/10.1007/s10734-015-9974-7.

Hammarfelt, B., & de Rijcke, S., (2014). "Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University", in Research Evaluation, 24(1), 63–77. http://doi.org/10.1093/reseval/rvu029.

Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). "The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities", in Research Evaluation, http://doi.org/10.1093/reseval/rvv040.

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

Hemlin, Sven, Barlebo Rasmussen, Søren, (2006) "The Shift in Academic Quality Control", in Science Technology Human Values, vol. 31, no. 2, 173-198.

Hemlin, Sven, (2009) "Peer Review Agreement or Peer Review Disagreement: Which Is Better?", Journal of Psychology of Science and Technology, Volume 2, Number 1.

Hicks, D. (2004). "The four literatures of social science", in H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), Handbook of quantitative science and technology research, Dordrecht: Kluwer Academic, 473–496.

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.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., Rafols, I., (2015). "The Leiden Manifesto for Research Metrics", Nature, vol. 520, 23 April 2015, p. 429-431

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.

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", in Research Evaluation, 22(5), 369–383. http://doi.org/10.1093/reseval/rvt008

Kekäle, J. (2002). "Conceptions of quality in four different disciplines", in Tertiary Education and Management, 8(1), 65–80, http://doi.org/10.1023/A:1017947005085.

Klein, J. T. (2008). "Theoretical perspectives on team science, Evaluation of Interdisciplinary and Transdisciplinary Research: A Literature Review", American Journal of Preventive Medicine, Volume 35, Issue 2, Supplement, August 2008, p. S116-S123.

Mali, F., Pustovrh, T., Platinovsek, R, Kronegger, L, & Ferligoj, A. (2016). "The effects of funding and co-authorship on research performance in a small scientific community", Science and Public Policy, online first. doi:10.1093/scipol/scw076.

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

Morris, M. (2008). Evaluation ethics for best practices: cases and commentaries, New Yonk, The Guilford Press.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., (2005). Handbook on constructing composite indicators. Methodology and user guide. Paris: OECD.

Nederhof, A. J. (2006). "Bibliometric monitoring of research performance in the social sciences and the humanities: A review", Scientometrics, 66(1), 81–100, http://doi.org/10.1007/s11192-006-0007-2.

Nussbaum, M. C. (2010). Not for Profit: Why Democracy Needs the Humanities. Princeton University Press: Princeton.

Ochsner, Michael, Hug, Sven, Daniel, Hans-Dieter, (2012). "Indicators for Research Quality in the Humanities: Opportunities and Limitations", Bibliometrie - Praxis Und Forschung, 1.

Ochsner, M., Hug, S. E., & Daniel, H.-D. (2014). "Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies." Zeitschrift für Erziehungswissenschaft, 17(6 Suppl.), 111–132. http://doi.org/10.1007/s11618-014-0576-4

Penfield, T., Baker, M.J., Scoble, R., Wykes, M.C. (2014). "Assessment, evaluations, and definitions of research impact: A review", in Research Evaluation, 23 (1), 21-32.

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

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

Trisha Greenhalgh, James Raftery, Steve Hanney and Matthew Glover, (2016) "Research impact: a narrative review", in BMC Medicine, 14:78.

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

Williams, G., & Galleron, I. (2016). "Bottom Up from the Bottom: A New Outlook on Research Evaluation for the SSH in France." In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), Research Assessment in the Humanities. Towards Criteria and Procedures (pp. 181–198). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-29016-4_14

AUTHORS

IOANA GALLERON
Paris III University, (France)
E: galleron@evalhum.eu

MICHAEL OCHSNER
ETH Zürich/ FORS, Lausanne (Switzerland)
E: ochsner@evalhum.eu

JACK SPAAPEN
KNAW (the Netherlands)
E: jack.spaapen@knaw.nl

GEOFFREY WILLIAMS
University of Grenoble-Alpes (France)
E: williams@evalhum.eu