

Tilburg University

Redesigning research evaluation practices for the Social Sciences and Humanities

de Jong, Stefan; Balaban, Corina; Holm, Jon; Spaapen, Jack

Published in:
Deeds and Days

DOI:
[10.7220/2335-8769.73.1](https://doi.org/10.7220/2335-8769.73.1)

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
de Jong, S., Balaban, C., Holm, J., & Spaapen, J. (2020). Redesigning research evaluation practices for the Social Sciences and Humanities: Perspectives from the European network for research evaluation in the social sciences and humanities (ENRESSH). *Deeds and Days*, 2020(73), 17-35. <https://doi.org/10.7220/2335-8769.73.1>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

REDESIGNING RESEARCH EVALUATION PRACTICES FOR THE SOCIAL SCIENCES AND HUMANITIES: PERSPECTIVES FROM THE EUROPEAN NETWORK FOR RESEARCH EVALUATION IN THE SOCIAL SCIENCES AND HUMANITIES (ENRESSH)



STEFAN DE JONG

Leiden University

CORINA BALABAN

The University of Manchester

JON HOLM

Norwegian Research Council

JACK SPAAPEN

Royal Netherlands Academy of Arts and Sciences

ISSN 1392-0588 (spausdintas)

ISSN 2335-8769 (internetinis)

<https://doi.org/10.7220/2335-8769.73.1>

2020. 73

This article is dedicated to the memory of Paul Benneworth, a driving force behind the ENRESSH network and leader of Work Group 2 – Relevance and Impact.

SUMMARY. This paper presents the findings of the ENRESSH network with relevance to academic and policy communities. As a recently completed COST action running between April 2016 and April 2020, ENRESSH consists of over 150 researchers and policy makers from 40 countries across Europe and beyond. Through its ongoing research and networking, ENRESSH has worked towards two goals: 1) to enhance the visibility of research in the social sciences and humanities (SSH), as well as its potential to address questions around major societal challenges and 2) to develop comprehensive evaluation methods that better fit how researchers in the SSH communicate. In response to the first goal, ENRESSH proposes that, in order to enhance the visibility of the SSH, impact assessments need to be more inclusive, in particular towards the values and practices of the SSH. This could be achieved by focusing on the interaction processes between researchers and stakeholders, by acknowledging different types of impact and impact pathways, and by improving the understanding of impact by SSH researchers. In response to the second goal of improving evaluation methods, ENRESSH recommends that (extended) peer review constitutes the basis for evaluation, given that it is more suitable to meet the specific cognitive, and often context-dependent, challenges of SSH research. While peer review may continue to be supported by quantitative measurements, these need to be aligned with the types of output and communication patterns prevalent in the SSH. Finally, in order to implement our research-based vision, the ENRESSH community advocates a closer collaboration between SSH researchers, policy makers and societal stakeholders.

KEYWORDS: social sciences, humanities, research evaluation, societal impact, internationalization.

PREAMBLE

ENRESSH is a European Cooperation on Science and Technology (COST) action (15137) that ran between April 2016 and April 2020. It consists of a network of over 150 researchers and policymakers from 40 countries in Europe and beyond. Members of ENRESSH share a professional interest in research evaluation and societal impact of the humanities and the social sciences. ENRESSH investigates the effects of current research evaluation systems in European countries from the perspective of research(ers) in the social sciences and humanities (SSH). Based on this investigation it provides concrete advice on how to make those systems work better to enhance both quality and relevance of SSH research. This article is a reflection on the results of research conducted in the context of ENRESSH and on the perspectives developed during other activities organized by the network, including its three training schools, eight work group meetings, and a number of stakeholder events.

INTRODUCTION

The ENRESSH network (www.ENRESSH.eu) emerged in 2016 from a European group of social sciences and humanities (SSH) researchers working in the EvalHum initiative. EvalHum is a Europe-wide initiative concerned with research evaluation, innovation and impact in the SSH. It brings together experts in research evaluation and research impact in the SSH, including institutional stakeholders and researchers in a wide variety of SSH disciplines (www.evalhum.eu). A key aim of EvalHum is to gain insights into questions regarding evaluation, especially in the humanities and social sciences.

The premise is that many SSH researchers feel uncomfortable with the fact that most evaluation systems appear to be built on the practices used in the science, technology, engineering and mathematics (STEM) fields. Here, communication about results of scientific research is commonly done via publications in international journals in English, and the quality of research is frequently measured using bibliometric indicators. SSH disciplines can be disadvantaged in evaluations that are designed based on STEM practices. For instance, research in the SSH is more likely to be conducted in local languages, engaging with national literature, culture, history or law, all of which have their own importance to local/ regional development. The relevance of SSH research is thus of a different nature, and may consist in strengthening democratic values and debates. A topical example is the black lives matter movement, including waves of iconoclasm that is spreading all over the world. How can we address this in a meaningful way without knowledge of

history, racial and cultural tensions and socio-economic differences, for instance? Therefore, SSH research needs to be recognized for its own merits and judged in a way that fits its own working and communication modes. ENRESSH proposes that policy-makers become more creative when designing evaluation procedures, in order to support the core values of the SSH.

Two issues need to be addressed to improve the situation. First, production and communication modes in the SSH should be leading the design of evaluation procedures and enable SSH research to be assessed on its own terms. Second, the impact of SSH research on societal challenges, which is currently under-recognized, could be publicized much more openly. Both issues have become major concerns in discussions among SSH scholars and also in policy discussions in individual countries and at European level.

The importance of these concerns is reflected in the fact that ENRESSH has gathered over 150 participants from 37 European countries, China, Mexico and South-Africa. They represent a wide variety of fields, such as economics, history, law, literature and languages, culture, migration, gender studies, and also fields outside the SSH such as environmental sciences, health studies and chemistry. Fields that study academic research and policy have also taken part, including science and technology studies (STS), scientometrics, information and communication sciences. In addition, the group of researchers was joined by policy makers and research administrators, all part of the ENRESSH network.

This variety of profiles has allowed for in-depth cross-national and transdisciplinary knowledge exchange about various disciplinary knowledge production modes, interests and practices in the context of possible approaches to evaluation of quality and impact of research in the social sciences and humanities (SSH). In these discussions, resulting in 20 peer-reviewed publications so far, sensitive topics were tackled, such as the pros and cons of bibliometric and non-bibliometric methods, the ratio between a focus on 'excellence' of research and its 'societal value', the tension between the points of view of evaluators and of those who were evaluated, and so on. These discussions support two main aims.

The first main aim of ENRESSH and its members is to enhance the visibility of SSH research and its potential to address questions around major societal challenges. Most often, when faced with such challenges, policy-makers direct their attention to the STEM fields. Yet, many of the challenges that they aim to address are situated within an SSH area of expertise. The current COVID-19 pandemic of course is a major medical challenge, but it has quickly become clear that it has major implications for our social, economic and cultural systems. Other examples are the migration crisis in Europe, the instability of financial and democratic systems, religious tensions in the Middle East, and so on. The input of the SSH

is further required also when tackling technical challenges. For instance, climate change, the energy transition, sustainability and health challenges include major behavioral, political and social questions, as well as technological advancements, which can only be successfully developed and implemented when analyzed alongside SSH research perspectives.

The second aim is to develop comprehensive evaluation methods that better fit how SSH researchers communicate with each other, with other fields of research, and with society at large. For ENRESSH, comprehensive means that stakeholders are involved not only in terms of collaboration in research projects, from initiation to dissemination, but also in the design of evaluation methods. This aim, which supports the first aim of enhancing the potential of SSH, is important for two reasons: (1) many of the evaluation methods in use today are still based on the way STEM fields produce and communicate their research results, including communication with immediate stakeholders which in the case of STEM often is situated in the private sector; and (2) most methods are still weak when it comes to assessing the contribution of research to societal questions. ENRESSH members are dedicated to addressing these two broad sets of research questions through conducting research, organizing meetings and workshops, and stimulating early career researchers to contribute to collaborative projects.

The remainder of this paper is organized as follows. In section 2 we discuss the alignment of research evaluation to policy goals. In section 3, we briefly summarize some major issues concerning evaluation of the SSH as identified by academic research. Section 4 argues that international and cross-disciplinary comparison, even within the SSH, poses further challenges for research evaluation. Sections 5 and 6 present the main achievements of ENRESSH in terms of enhancing the visibility of the SSH contribution to tackling societal challenges and improving evaluation methods for the SSH. We conclude with section 7, which includes a reflection on our results and directions for future research and policy making on the evaluation of the social sciences and humanities.

POOR ALIGNMENT OF RESEARCH EVALUATIONS TO POLICY GOALS

Arguably the main rationale behind research evaluations is an economic one: when public money is allocated, it is necessary to ensure that funding serves science and society in the best possible way. For that, it is important to know where the best quality is, and whether the research will have an impact both on the scientific community and on society. Other rationales range from selecting individuals for academic positions (career related evaluations) to improving institutional strength

(research policy and management evaluations) to setting priorities for national and supranational (EU) systems (funding evaluations).

Evaluations also serve many different purposes that can be divided into two large groups: summative evaluations used for the (re)distribution of funding and formative evaluations used for improvement of research practices and institutional learning (Scriven 1996). Summative evaluations are often interested in the attribution of results and effects, whereas an interest in processes and contributions is usually associated with formative evaluations (Spaapen, Van Drooge 2011). Hence, different goals of evaluations call for different foci and types of evidence that support the evaluation.

Yet, there is a growing concern in the communities of both (supra) national policymakers and academic institutions that the way research is evaluated - both for distribution of funding and for learning - is not in line with established policy goals and ambitions, nor with developments in the academic research world itself. Decisions about the allocation of public funds and the increasingly pressing demands from society call for an informed consideration of many different interests, which are not necessarily included in current evaluation practices, and instruments that are able to address complex question (Wernli et al. 2016).

As early as 2010, the European Commission called for caution in research evaluations, by stating that:

Different publication and dissemination practices characteristic of different disciplines and fields can be positively and negatively affected by the choice of indicators.

The European University Association (2019) also recently recognized that the current selection of indicators is skewed:

While university missions concentrate on education, research and innovation, current incentive and reward structures predominantly focus on research output.

It also specifies the effect of selecting indicators that only capture a limited range of academic activities:

Research assessment [...] suffers from a growing mismatch between what society and the academic community value and what is incentivised and rewarded.

Hence, it concludes that “There is no single set of indicators capable of capturing the complexity of research and research assessment.”

The above mentioned concerns have led policy makers to initiate a number of projects to map and address them. Evaluation systems across Europe have been documented in earlier studies of research evaluation, for example, see *Assessing Europe’s University-Based Research* (Expert Group on Assessment of University-Based Research, 2010) and *MLE on Performance-based Research Funding Systems of Public Research Organisations in 2017* (European Commission, 2020),

both initiated by the European Commission. The Open Science agenda has recently provided a new perspective on the role of evaluation systems in the implementation of research policies; see for example the European Commission's expert group on Next-generation metrics: Responsible metrics (Wilsdon et al., 2017), the European University Association's Roadmap on Research Assessment in the Transition to Open Science (European Association of Universities, 2018) and Briefing on Reflections on University Research Assessment (Saenen, Borell-Damián, 2019). The European Commission's expert group on Altmetrics (Wilsdon et al., 2017) has called for the widening the attention of research evaluation to include social media, blogs, webinars and citizen science projects. Another example is the Horizon 2020 project New Horizon (<https://newhorizon.eu/>), which aims to further integrate responsible research and innovation in research and innovation systems, with the ultimate goal of bridging the gaps between these systems and society at large.

ADVERSE EFFECTS OF CURRENT EVALUATION PRACTICES ON THE SCIENCE SYSTEM

Over the past decade, scholars have identified a number of problematic issues in current research evaluation practices. De Rijcke et al. (2016) reviewed the evaluation literature and concluded that current evaluation practices induce goal displacement and affect strategic behavior among academics and academic organizations. This affects the research system on at least four different levels.

On the level of academics, this leads to a transformation of research practices. This includes changes in research agendas, which decrease research diversity (e.g., Whitley, 2007), and changes in dissemination patterns, caused by a dominant focus on journal publications and this leads to homogenization and a lack of societal orientation of researchers (e.g., Willmott, 2011). When national level criteria are applied to the individual level, academics can experience a lack of agency, as Burrows (2012) demonstrates for the case of the H-index.

On the level of research domains, a focus on international journal publications does not do justice to the diverse modes of communication in the social sciences and humanities, including books, national language journals and non-scholarly publications (Hicks, 2004; Sivertsen, 2016; Spaapen, Van Drooge, 2011). Furthermore, evaluation practices result in tensions for interdisciplinary research, as there is a lack of consensus on how to evaluate this type of research (Belcher et al., 2016; Spaapen et al., 2020). The development of relevant and meaningful benchmarks and indicators however is not yet part of evaluation systems and procedures. When considering impacts on society, a focus on economic impacts fails to showcase the

full spectrum of societal benefits of the social sciences and humanities (Gibson, Hazelkorn, 2017), and arguably other domains as well.

On the level of organizations, it may contribute to a transfer market for academics whose work positively contributes to evaluation outcomes and thereby amplifies quality differences between universities, as the Research Assessment Exercise caused in the UK as early as the 1990s (HEFCE, 1997, cited by Barnard, 1998; Colwel et al, 2012), resulting in a tighter hierarchical link between universities and governments. For example, universities have started to implement local evaluation systems, that to a large extent mimic the criteria that national systems include (Hammarfelt et al., 2016). Once more, this lowers the diversity of research practices.

On the international level, the focus on certain indicators, most often publications in international peer-reviewed journals, especially top journals, negatively affects opportunities for contextualised research which addresses local societal issues in collaboration with non-academic societal actors (Bianco et al., 2016). In addition, the bias against languages other than English is particularly harmful to the SSH. Finally, SSH evaluation is still significantly impeded by the lack of robust and valid data on publications and societal impact. Although data is currently being collected (through project evaluation, programme evaluation, institution evaluation, etc.), it is neither harmonized nor complete at the European level.

Besides the above mentioned adverse effects, some authors question the meaningfulness of current indicators altogether. It is unclear whether these traditional indicators, based on publications and citations have ever been able to capture scientific progress and performance (Reale et al., 2018). And, if they ever did, some argue that their uncritical use leads to biased and even incorrect results (Barré, 2019).

INCREASING GAP BETWEEN EVALUATIONS AND PRACTICES

A major activity of ENRESSH has been to compare evaluation practices in different countries and deepen the understanding of existing tensions. Thus, we are now better equipped to enhance the visibility of SSH research and its potential to address societal questions, as well as to improve evaluation methods for this domain. International comparisons are at the core of our work, given the international composition of our network. This section describes our most significant contributions regarding the tensions in current evaluation practices.

An analysis of publication patterns in the SSH across eight European countries and three disciplines by Kulzyczki et al. (2018) shows that there are large differences in publication patterns, in terms of language, between countries and disciplines. In Western European countries, publications in English may account for up to 80%

of the total number of publications, as is the case in Belgium, whereas in Eastern European countries this share might not exceed 20%, as is the case in Poland. Still, in Belgium, law scholars publish around 40% of their work in English, while their colleagues in Economics and Business publish around 90% of their work in English. Even within the same discipline in different countries there are significant differences, meaning, for example, that the publication profile of the average theology scholars in Poland looks quite different from the average theology scholar in the Belgium. All in all, this study suggests that evaluations across countries and disciplines should provide a generic framework that allows for comparison while at the same time respecting these differences.

Although the role of book publishing in evaluations is small compared to the role of journal publications, publishing in books traditionally constitutes a large share of publications in the SSH. Popular belief suggests that publishing in books is declining. ENRESSH aims for an evidence based discussion and has investigated several types of book publishing, such as edited books, chapters, monographs and textbooks. For instance, Giménez-Toledo et al. (2019) compared how scholarly books are taken into account in research evaluation in 19 European countries. They identified four main modes of including book publications, varying from non-formalized systems which rely less on quantitative indicators, to systems that use supra-institutional databases, quality labels for publishers or rankers, or a combination, for comparisons. They found that researchers in different countries are being assessed and incentivised in different ways. Similarly, Engels et al. (2018) studied variations in book publishing in the humanities across five European states/ regions: Finland, Flanders, Norway, Poland and Slovenia and across a variety of disciplines. They concluded that in these countries and disciplines book publishing is not declining. Yet, the share of book publishing differs per country and discipline. Again, this suggests the importance of a generic framework that can be adapted to different contexts.

A similar picture of differences between countries emerges when looking at societal impact practices across Europe. De Jong and Muhonen (2020) analyzed 60 SSH impact case studies that ENRESSH members conducted in 17 European countries. They considered the motivation of scholars to have an impact on society, the stakeholders they collaborated with, the way stakeholders were involved in the research and how results were disseminated. Although motivations are similar for Eastern and Western European scholars, Western European scholars collaborate with a larger variety of stakeholders, tend to be involved in co-creation more often and use a wider range of dissemination channels, including documentaries, consultancy and master classes for professionals. On top of that, impact reporting styles tend to be more specific and detailed in Western European countries, possibly because these scholars are more accustomed to an impact culture. When regarding

these differences in the light of European level evaluation criteria for impact, which emphasize the effectiveness of methods proposed for stakeholder interaction, it becomes clear that Eastern European scholars are at a disadvantage.

In practice, the above mentioned differences are hardly considered in evaluation practices. The development of national evaluation systems has been a defining feature of research policy in Europe over the last decades, responding to increased globalization of research, and often aimed at strengthening the international competitiveness of the research environment in a given country. Ochsner et al. (2018) made an inventory of 32 national evaluation systems in Europe. They found that many systems only superficially consider differences between research production and communication in diverse disciplines and fields. Given the concern for international competitiveness, SSH-research in particular is affected. In many countries, national evaluation systems put pressure on the SSH to adapt its scholarly practices to those of the STEM-fields, making publication of research articles in international journals the standard format of scholarly communication. Accordingly, we observe a standardization of evaluation criteria modelled on the STEM fields creating evaluation procedures that are ill-adapted or even inappropriate for SSH research paradigms. When focusing on impact, a similar lack of SSH sensitivity can be observed on the European level. Impact on private companies is given particular attention in impact evaluations, but impact on public organizations is not specifically included in the criteria. Although collaboration with companies is familiar to SSH scholars, their primary stakeholders are found in the public sphere (de Jong, Muhonen, 2020).

The contribution of ENRESSH to the analysis of tensions in current evaluation practices is its unprecedented breadth in terms of international comparison – comprising input from 37 European countries. Through this comparison, we learn that academic practices are even more varied across countries and disciplines than we already knew, and that these differences are taken even less into account than anticipated. A one-size-fits all evaluation system focused on a particular notion of research excellence risks reducing institutional diversity by ignoring diversity in institutional goals and roles, as the League of European Research Universities (Van den Akker, Spaapen, 2017) emphasizes as well.

A BETTER UNDERSTANDING OF IMPACT IS KEY TO ENHANCING THE VISIBILITY OF SSH CONTRIBUTIONS TO SOCIETY

Key to enhancing SSH visibility is a better understanding of and attention to the processes that lead to its contributions to society. Several ENRESSH studies have

considered these processes, and while their primary focus is on the SSH, their insights are valuable to other domains as well. We discuss these studies in this section.

Sivertsen and Meijer (2020) argue for more attention to ‘normal impact’ rather than ‘extraordinary impacts’. They define normal impact as “the results of active, productive, and responsible interactions between (units of) research organizations and other organizations according to their purposes and aims.” They stress that these impacts often occur informally on the individual or research group level, but that they may be the result of formal structures on the organizational level as well. Extraordinary impact is defined as “more rare incidences where traditional and typical or new and untypical interactions between science and society have unexpected widespread positive or negative implications for society.” According to Sivertsen and Meijer, current evaluation practices focus primarily on extraordinary impacts. As such, only a few instances of societal impacts and efforts of researchers are being made visible. They suggest that by focusing on normal impacts as well, the breadth of societal contributions and corresponding efforts can become more visible.

Muhonen, Benneworth and Olmos-Penuela (2020) also focus on impact creation processes. They analyzed 60 impact case studies from 51 SSH disciplines and 16 countries across Europe to identify the mechanisms that generate impact. A total of sixteen mechanisms were identified, which can be grouped into four main categories. The first is dissemination, characterized by scientific progress preceding societal progress. The second is co-creation, characterized by simultaneous changes in science and society. The third is reacting to societal change, characterized by alignment of science to societal changes. The final category is driving societal change, characterized by changes in disciplinary directions to address societal needs more proactively (note that the many actors and interactions involved make it extremely hard to attribute impacts to specific researchers or interactions). If impact evaluations aim to capture all research impacts, the potential value of all sixteen mechanisms should be recognized.

Boshoff and De Jong (2020) took a complementary quantitative approach and analyzed survey responses of 485 researchers from 46 countries across sub-Saharan Africa – to include a broader geographical perspective on impact. Respondents were asked to assess five stylized examples of research impact from the SSH, STEM and biomedical domains, and from agricultural and interdisciplinary research. They show that among researchers from all domains there are differences in terms of what they understand to be the societal impact of academic research. Some only consider ultimate improvements to society, such as improved food security, as impact, whereas others also consider research products, such as a new crop variety, as impact. Most notably, a stylized example of impact from the humanities was

least recognized as societal impact, even by researchers from the social sciences and humanities. These findings highlight the importance of a better and possibly shared understanding of societal impact by academics with special attention given to the abilities of the SSH to improve the understanding of impacts from the SSH. If academics do not recognize their own or each other's impact in evaluative contexts, impacts are less likely to become visible. De Jong and Muhonen (2020) found that reporting on impacts is not clear-cut for all academics, emphasizing the importance of understanding what impact is and how to report it for making impact visible as well.

Referring to the Management of Social Transformation (MOST) programme of UNESCO, Sigurdarson (2020) discusses the risk of goal displacement induced by STEM oriented evaluation systems. The UNESCO report states that managing social transformation is not only about technical solutions; it is also about imagining creative solutions, in which the humanities have a key role to play (cited by Sigurdarson: 72). The author reviews this role of the humanities in terms of its ability to strengthen academic and other communities, enabling them to handle social change. Interestingly enough in the current times of a pandemic, Sigurdarson refers to an example given by Werkheiser (2016) from the field of epidemiology, an interdisciplinary field between medicine and SSH. The example illustrates that it is not only about the capacity of communities to adopt the knowledge of experts, but also about the ability of communities to contribute in certain areas (Werkheiser: 40, cited by Sigurdarson: 73).

All in all, these studies conclude that to enhance the visibility of the contributions that the SSH make to societal progress, impact assessments need to be more inclusive, in particular towards SSH practices. This could be done by focusing on processes of interaction between researchers and stakeholders, and by being sensitive to different types of impacts and pathways. Such a focus includes assessing the commitment of researchers to societal impact and the choices that they make in these processes. For example, scholars working on urgent societal issues and/or those who are involved in long-term co-creation with stakeholders made an active decision to do so (ENRESSH, n.d.). A second promising avenue is to improve the understanding of scholars of these processes, as they often are responsible for either reviewing or submitting information for reviews. If scholars do not fully recognize the breadth and quality of impact generating processes and impacts, they will not be able to properly demonstrate and reward them. In other words, improving scholars' understanding of impact is key to capture all impacts in research evaluations.

IMPROVING EVALUATION METHODS FOR THE SSH

Despite the above mentioned challenges, some systems do allow for variety, for instance, by adapting the metrics in use, widening the peer review system, or via bottom-up procedures that differ per institution. ENRESSH concludes that for SSH research, (extended¹) peer review should always be the basis for evaluation, possibly supported by quantitative measurements as long as they fit output and communication patterns common in the SSH. Peer review can meet the specific cognitive challenges that come with SSH research, such as the context dependency of much of SSH research, conflicting research paradigms, the diversity of publication outputs and the specific importance of books or monographs, the importance of local languages, interdisciplinarity, and the relation to the Open Science agenda. Of course, challenges of peer review, such as the risk of gender bias, conservative bias, and the workload for all parties involved, also need to be considered.

All in all, with the exception of the British Research Excellence Framework, the Dutch Strategy Evaluation Protocol (SEP), and a handful of other peer review based systems for distribution of institutional funding, most of the Performance-based Research Funding Systems (PRFS) in Europe are indicator based².

The latest version of the Dutch SEP system, which will be implemented in 2021, looks particularly promising for the SSH and the evaluation of societal impact. Not only because the SEP has implemented some ideas about appropriate evaluation methods for the humanities that were developed in the QRiH project (<https://www.qrih.nl/en>), but also because one of its central targets is now the research strategy of university departments. This suggests that looking forward and qualitative assessment are now at least as important as looking backwards and using performance indicators. Of particular importance here is the assessment of a research unit in light of its own aims and strategy, based on a narrative with supporting evidence provided by the unit.

What the SEP tries to accomplish is the attuning of the different demands that academic researchers have to meet. As Van den Akker and Spaapen (2017) argue, universities are at the crossroads of international academic competition and local, national or European policy demands. Research has to be excellent in an international context, and at the same time relevant to challenges of the society in which it is grounded. For current evaluation systems, the challenge is to find ways to assess and value both aspects.

¹ Extended peer review explicitly includes perspectives of experts from society, such as stakeholders, and/or other disciplines in assessment procedures.

² MLE on Performance-based Funding of Public Research Organisations (EC 2017).

Overall, the main weaknesses of evaluation systems with regard to SSH research are the following:

- Indicators used in PRFS are often based on international bibliometric databases such as Web of Science or Scopus with poor coverage of non-English journals and books formats that are important outlets for SSH research.
- The diversity of publication patterns reflects the diverse roles of SSH research in society. A streamlining of SSH research publications into journal article publishing in English may diminish the public value of SSH and impede its societal impact.

Evaluations focused on learning will typically be based on peer review where experts assess performance based on a mix of quantitative and qualitative data. The main challenges of these systems with regard to SSH research are the following:

- The roles of SSH for education and society at large are often overlooked in evaluations of university research.
- Evaluation criteria are often modelled on STEM-fields.
- Disciplines of specific national importance (language, literature, history) are at risk of being underrated in internationally oriented evaluations performed by international peers.

Reliable and comparable data is crucial for research assessment frameworks to produce meaningful insights. Well-known databases such as Web of Science and Scopus are known for their poor coverage of SSH outputs and languages other than English. Due to the differences in national publication information systems and the resulting poor interoperability of these systems, international comparison is further complicated. A European level database could overcome this problem. ENRESSH took up this challenge and developed the VIRT-A-ENRESSH Proof of Concept (PoC) Database. The technical details of the database are explained in Puuska et al. (2018). This database includes institutional data from six universities across four European countries: the University of Helsinki, the University of Jyväskylä and Tampere University of Technology in Finland, the University of Antwerp in Belgium, the University of Oslo in Norway and the University of Carlos III Madrid in Spain. The integration of existing institutional data required a limited investment and resulted in a cross-national operable database including 50,000 references. As European scholars often publish in languages other than English, the expansion of this database will result in an important asset for improving evaluations in the SSH. Such a database would include all types of publications and possibly other research outputs. As participation is open for all European countries, over time, the database could lead to improved coverage and international comparison of research outputs in the SSH.

CONCLUDING REMARKS

One could say that the main implicit aim of ENRESSH is to tell the world how interesting, important and valuable the work of SSH scholars is. This has translated into two main concrete aims: (1) to enhance the visibility and usefulness of the SSH in order to improve its capacity to help tackle societal challenges, and (2) to reform evaluation methods to better fit the research and communication practices of the SSH.

From the start, ENRESSH has studied researchers operating at all organizational levels: the inter-personal level, the institutional level (where researchers operate in the context of rules and regulations, as well as competition with other academic fields), and the systemic level, referring to the larger science and innovation system, with national and international ramifications. Three key themes run throughout most of the findings of ENRESSH at these levels: translation, diversity of national contexts, and internationalization.

Regarding translation, researchers need to communicate their aims and findings at different levels, and do so in a way that bridges the existing variation in interests and understandings. But this is one side of the medal. The other side is that stakeholders outside the realm of SSH research, whether they are researchers from other disciplines, policy-makers or other societal parties, to understand SSH disciplines and communicate with them. ENRESSH finds that most of the misalignment between research goals and policy goals is due to poor translation and unmediated communication.

A second theme is the diversity of national contexts. ENRESSH finds that the practices of evaluation, ideas about impact and careers are deeply embedded in national contexts. These differences affect the emphasis that SSH receives in evaluation practice, or societal impact for that matter. In some cases, this is in response to specific national needs; in other cases, this is a matter of tradition and inertia. While it is important to appreciate differences between countries, there are many similarities as well. This should make the process of attuning feasible.

Researchers and their organizations operate in an international and multi-dimensional environment in which they are exposed to different ways of working other than their own. To respect local contexts while simultaneously acknowledging internationalization, including its benefits for research, ENRESSH concludes that an overarching framework for evaluation and impact is necessary, provided it offers a common approach, while at the same time respecting national and disciplinary variety. The relationship between the overarching framework and local contexts is, of course, a complex one. It is not only research ideas that travel, but also policies on research, which also follow international trends. Ideally, such an overarching system

would be a combination of bottom up input (from the academic community) and top down framework (initiated by policy makers) allowing for institutional and national variety, while at the same time coordinating societal demand.

Universities in Europe and beyond are at a crossroad. On the one hand, they need to engage in international competition, which has been actively promoted in research policies throughout the last decade. Many countries adopt the scholarly practices of the STEM-fields, making publication of research articles in international journals the standard format of scholarly communication. However, the STEM focus of research policies has been detrimental to the visibility and valuation of SSH research.

On the other hand, there has been a growing demand for researchers to be relevant to society, from engaging with relatively small and local issues to addressing the grand societal challenges, or the global UN sustainable development goals. Although SSH research can make many meaningful contributions here, the development of broadly accepted evaluation systems for impact is still lagging behind.

The grand rationale for adapting evaluation systems to become more appropriate for SSH research is that SSH research underpins and strengthens democracy and for that important reason alone needs to be recognized for its own merits in teaching and developing critical thought. Additionally, SSH research contributes to a considerable extent to the understanding of many societal challenges, such as the current health and economic crises, issues of (global) migration or other conflicts arising from religious, cultural and socio-economic differences. The ability of SSH researchers and their institutions to respond to these pressing societal issues are influenced by the evaluation systems constructed by policy-makers. This is why we are convinced that it is in the interest of policy-makers to better understand how these evaluation systems affect SSH research in order to develop evaluation procedures that are more inclusive of SSH research and communication practices. Following the results of the ENRESSH work, it is important to consider the following issues:

- ENRESSH has shown that in the realm of SSH research, different factors need to be considered when setting up research evaluations. For example, one needs to reflect on differences in the use of foreign languages when publishing, the kind of medium used for publication (books versus articles), or the use of other media to communicate (for example the use of social media, or audio-visual communication). Therefore, evaluations should provide generic frameworks that allow for comparison while at the same time respect these differences.
- Similarly, when it comes to implementing societal impact in evaluations, those who design them need to consider that impact is highly contextual. Different societies need different things. Moreover, ENRESSH finds that

although motivations are similar for Eastern and Western European scholars, Western European scholars collaborate with a larger variety of stakeholders, tend to be involved in co-creation more often and use a wider range of dissemination channels, including documentaries, consultancy and master classes for professionals (De Jong & Muhonen, 2020). In developing better ways of assessing impact, it is recommendable to focus not only on extraordinary impact (the remarkable examples that are the exception rather than the norm) but also on normal impacts, which include the full breadth of societal interactions and contributions. ENRESSH distinguishes sixteen categories of mechanisms to consider when assessing impact. These can be grouped into four main categories: dissemination, co-creation, the alignment of science and societal changes, and drivers of societal change.

- ENRESSH found that in most European countries, differences between research production and communication in diverse disciplines and fields do not – or only superficially – play a role, and that this affects the SSH much more than the STEM fields. This is evident, for instance, in the dominance of publications in international journals as an indicator of quality. An SSH friendly evaluation system should make serious room for other indicators, for example, those based on book publications. The development of reliable databases for this goal is important and necessary. ENRESSH took up this challenge and developed the VIRTIA-ENRESSH Proof of Concept (PoC) Database, which includes integrated institutional data from six universities across four European countries, resulting in a cross-national operable database including 50,000 references. Given the promising results of the PoC Database, which provides a generic framework, while allowing for national and institutional differences, we welcome research institutions to benefit from that database.

To sum up, it is an absolute necessity that, on the one hand, SSH researchers become much more assertive about their contributions to societal issues, whether these are on a small, local scale, or address bigger challenges in society; and on the other hand, policy makers create flexible procedures to do justice to the value and relevance of SSH research.

ACKNOWLEDGEMENTS

This article is based on work from COST Action 15137 ENRESSH, supported by COST (European Cooperation in Science and Technology). The authors wish to thank all members of ENRESSH for their valuable contributions to the network. In particular, they wish to express their gratitude to Emanuel Kulczycki, chair of

ENRESSH, and his predecessor Ioana Galleron, as well as group leaders Michael Ochsner, Tim Engels, Geoffrey Williams and Jolanta Šinkūnienė for coordinating the work of ENRESSH.

LITERATURE

- Barnard, J. W. Reflections on Britain's Research Assessment Exercise. *Journal of Legal Education*, 1998, vol. 48, no. 4, 467–495.
- Barré, R. Les indicateurs sont morts, vive les indicateurs! Towards a political economy of S&T indicators: A critical overview of the past 35 years. *Research Evaluation*, 2019, vol. 28, no. 1, 2–6; <https://doi.org/10.1093/reseval/rvy029>.
- Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., and Zornes, D. A. Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, 2016, vol. 25, no. 1, 1–17; <https://doi.org/10.1093/reseval/rvv025>.
- Bianco, M., Gras, N., and Surtz, J. Academic Evaluation: Universal Instrument? Tool for Development? *Minerva*, 2016, vol. 54, no. 4, 399–421; <https://doi.org/10.1007/s11024-016-9306-9>.
- Boshoff, N., and de Jong, S. P. L. Conceptualizing the societal impact of research in terms of elements of logic models: A survey of researchers in sub-Saharan Africa. *Research Evaluation*, vol. 2020, 29, no. 1, 48–65; <https://doi.org/10.1093/reseval/rvz020>.
- Burrows, R. Living with the h - index? Metric Assemblages in the Contemporary Academy. *The Sociological Review*, vol. 60, no. 2, 355–72; <https://doi.org/10.1111/j.1467-954X.2012.02077.x>.
- De Jong, S. P. L., and Muhonen, R. Who benefits from ex ante societal impact evaluation in the European funding arena? A cross-country comparison of societal impact capacity in the social sciences and humanities. *Research Evaluation*, 2020, vol. 29, no. 1, 22–33; <https://doi.org/10.1093/reseval/rvy036>.
- Engels, T. C. E., Istenič Starčič, A., Kulczycki, E., Pölönen, J., and Sivertsen, G. Are book publications disappearing from scholarly communication in the social sciences and humanities? *Aslib Journal of Information Management*, 2018, vol. 70, no. 6, 592–607; <https://doi.org/10.1108/AJIM-05-2018-0127>.
- ENRESSH. *Guidelines for evidence-led evaluation of research impact in the SSH*. ENRESSH, n. d.; <https://enressh.eu/wp-content/uploads/2020/04/Guidelines-for-evidence-based-evaluation-of-research-impact.pdf>.
- European Association of Universities. *EUA Roadmap on Research Assessment in the Transition to Open Science*. Brussels: European Association of Universities, 2018; <https://eua.eu/resources/publications/316:eua-roadmap-on-research-assessment-in-the-transition-to-open-science.html>.
- European Commission. *MLE on Performance-based Funding of Public Research Organisations*. Brussels: European Commission, 2020; <https://rio.jrc.ec.europa.eu/policy-support-facility/mle-performance-based-funding-public-research-organisations>.
- Expert Group on Assessment of University-Based Research. *Assessing Europe's University-Based Research*. Brussels: European Commission, 2010; https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf.
- Gibson, A. G., and Hazelkorn, E. Arts and humanities research, redefining public benefit, and research prioritization in Ireland. *Research Evaluation*, 2017, vol. 26, no. 3, 199–210; <https://doi.org/10.1093/reseval/rvx012>.
- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Guns, R., Kulczycki, E., Ochsner, M., Pölönen, J., Sivertsen, G., and Zuccala, A. A. Taking scholarly books into account, part II: A comparison of 19 European countries in evaluation and funding. *Scientometrics*, 2019, vol. 118, no. 1, 233–251; <https://doi.org/10.1007/s11192-018-2956-7>.
- Hammarfelt, B., Nelhans, G., Eklund, P., and Åström, F. The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 2016, vol. 25, no. 3, 292–305; <https://doi.org/10.1093/reseval/rvv040>.
- HEFCE (Higher Education Funding Council of England). *The Impact of the 1992 Research Assessment Exercise on Higher Education Institutions in England*. London, UK: HEFCE, 1997.
- Hicks, D. The Four Literatures of Social Science. In *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. Edited by H. F. Moed,

- W. Glänzel, & U. Schmoch. Dordrecht: Springer Netherlands, 2004, 473–496; https://doi.org/10.1007/1-4020-2755-9_22.
- Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Starčić, A. I., and Zuccala, A. Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics*, 2018, vol. 116, no. 1, 463–486; <https://doi.org/10.1007/s11192-018-2711-0>.
- Muhonen, R., Benneworth, P., and Olmos-Peñuela, J. From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. *Research Evaluation*, 2020, vol. 29, no. 1, 34–47; <https://doi.org/10.1093/reseval/rvz003>.
- Ochsner, M., Kulczycki, E., and Gedutis, A. The Diversity of European Research Evaluation Systems [Article in monograph or in proceedings]. 1235; Centre for Science and Technology Studies (CWTS), 2018; <https://openaccess.leidenuniv.nl/handle/1887/65217>.
- Puuska, H.-M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., and Engels, T. C. E. *Proof of concept of a European database for social sciences and humanities publications: Description of the VIRTIA-ENRESSH pilot*. ENRESSH, 2018; <https://repository.uantwerpen.be/docman/irua/44ceb4/150108.pdf>.
- Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, B., Mosoni-Fried, J., Oliver, E., Primeri, E., Puigvert, L., Scharnhorst, A., Schubert, A., Soler, M., Soòs, S., Sordé, T., Travis, C., and Van Horik, R. A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. *Research Evaluation*, 2018, vol. 27, no. 4, 298–308; <https://doi.org/10.1093/reseval/rvx025>.
- Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., and Hammarfelt, B. Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, 2016, vol. 25, no. 2, 161–169; <https://doi.org/10.1093/reseval/rvv038>.
- Saenen, B., and Borell-Damián, L. *Reflections on University Research Assessment: Key concepts, issues and actors*. European Association of Universities, 2019; <https://eua.eu/resources/publications/825-reflections-on-university-research-assessment-key-concepts-issues-and-actors.html>.
- Scriven, M. Types of evaluation and types of evaluator. *Evaluation Practice*, 1996, vol. 17, no. 2, 151–161; [https://doi.org/10.1016/S0886-1633\(96\)90020-3](https://doi.org/10.1016/S0886-1633(96)90020-3).
- Sigurdarson, E. S. Capacities, capabilities, and the societal impact of the humanities. *Research Evaluation*, 2020, vol. 29, no. 1, 71–76; <https://doi.org/10.1093/reseval/rvz031>.
- Sivertsen, G. Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 2016, vol. 107, no. 2, 357–368; <https://doi.org/10.1007/s11192-016-1845-1>.
- Sivertsen, G., and Meijer, I. Normal versus extraordinary societal impact: How to understand, evaluate, and improve research activities in their relations to society? *Research Evaluation*, 2020, vol. 29, no. 1, 66–70; <https://doi.org/10.1093/reseval/rvz032>.
- Spaapen, J., and van Drooge, L. Introducing ‘productive interactions’ in social impact assessment. *Research Evaluation*, 2011, vol. 20, no. 3, 211–218; <https://doi.org/10.3152/095820211X12941371876742>.
- Spaapen, J., Vienni Baptista, B., Buchner, A., and Pohl, C. *Report on Survey among interdisciplinary and transdisciplinary researchers and post-survey interviews with policy stakeholders*. SHAPE ID, 2020; <https://www.shapeid.eu>.
- Van den Akker, W., and Spaapen, J. *Productive interactions: Societal impact of academic research in the knowledge society*. League of European Research Universities, 2017; <https://www.leru.org/publications/productive-interactions-societal-impact-of-academic-research-in-the-knowledge-society>.
- Werkheiser, I. Community Epistemic Capacity. *Social Epistemology*, 2016, vol. 30, no. 1, 25–44; <https://doi.org/10.1080/02691728.2014.971911>.
- Wernli, D., Darbellay, F., and Maes, K. *Interdisciplinarity and the 21st century research intensive universities*. League of European Research Universities, 2016.
- Whitley, R. Changing Governance of the Public Sciences. In *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*. Edited by R. Whitley & J. Gläser, Dordrecht: Springer Netherlands, 3–27; https://doi.org/10.1007/978-1-4020-6746-4_1.
- Willmott, H. Journal list fetishism and the perversion of scholarship: Reactivity and the ABS list. *Organization*, 2011, vol. 18, no. 4, 429–442; <https://doi.org/10.1177/1350508411403532>.
- Wildon, D., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., and Wouters, P. *Next-generation metrics: Responsible metrics and evaluation for open science*. European Commission, 2017.
- Wróblewska, M. *Impact evaluation in Norway and in the UK*. A comparative study, based on REF 2014 and Humeval 2015–2017, 2019. ENRESSH Working Paper 1.

Stefan de Jong, Corina Balaban, Jon Holm, Jack Spaapen

HUMANITARINIŲ IR SOCIALINIŲ MOKSLINIŲ TYRIMŲ VERTINIMO PERTVARKA:
ENRESSH (EUROPOS HUMANITARINIŲ IR SOCIALINIŲ MOKSLŲ VERTINIMO TINKLO)
POŽIŪRIS

SANTRAUKA. Straipsnyje akademei bendruomenei ir mokslo politikams pristatomi ENRESSH tinklo tyrimų rezultatai. ENRESSH yra COST veikla, trukusi nuo 2016 iki 2020 metų balandžio mėnesio, joje dalyvavo daugiau kaip 150 mokslininkų ir mokslo politikų iš 40 Europos ir kitų šalių. Iki šiol atlikti moksliniai tyrimai paskelbti 20 recenzuotų publikacijų ir sukurtas bendradarbiavimo tinklas. ENRESSH siekė dviejų tikslų: 1) padaryti HSM geriau matomus, taip pat parodyti šių mokslų potencialą sprendžiant socialinius iššūkius, 2) pasiūlyti visapusiškus vertinimo metodus, kurie geriau atitiktų HSM esmę. Norint pagerinti HSM matomumą, jų poveikio vertinimas turėtų būti didesnės apimtys. Tą galima padaryti susitelkiant į HSM tyrėjų ir jų socialinių partnerių bendravimą, įvairias galimybes daryti poveikį visuomenei ir patiems HSM mokslininkams geriau suprantant to poveikio reikšmę. Siekiant pagerinti vertinimo metodus, mokslininkų tarpusavio vertinimas (*peer-review*) turėtų būti vertinimo pagrindas, nes tik jo metu galima teisingai įvertinti kognityvinius, dažnai nuo konteksto priklausančius HSM ypatumus. Tarpusavio vertinimą gali papildyti kiekybiniai rodikliai, jei tik jie atitinka HSM publikavimo ir mokslinės komunikacijos ypatumus. Kad ENRESSH išvalgos būtų įgyvendintos, reikia esminio dalyko – HSM mokslininkų, mokslo politikų bei socialinių partnerių bendradarbiavimo ir savitarpio supratimo.

RAKTAŽODŽIAI: humanitariniai ir socialiniai mokslai, mokslinių tyrimų vertinimas, socialinis poveikis, tarptautiškėjimas.