

CHALLENGING EVALUATION IN SSH: ROUND TABLE DISCUSSION



ISSN 1392-0588 (spausdintas)
ISSN 2335-8769 (internetinis)
<https://doi.org/10.7220/2335-8769.73>
2020. 73

ROUND TABLE DISCUSSION PARTICIPANTS:

Paul Benneworth (Western Norway University of Applied Sciences) unexpectedly passed away on May 12, 2020. Paul was Professor of Innovation and Regional Development at the Western Norway University of Applied Sciences and a Senior Researcher at CHEPS (Twente University), the Netherlands. Within ENRESSH COST Action Paul was the leader of work group 2, which focused on societal impact and relevance of SSH research. Paul's research interests were related to the relationships between universities and societal change with a particular focus on social sciences and humanities research, as well as societal change in old industrial regions. Paul had undertaken a wide range of basic and applied research activities for a variety of funders including research councils, HE funding councils, the OECD, government departments across Europe, the European Commission and a number of regional authorities. Paul was the inspirational leader of the work group focusing on societal impact and relevance of SSH research at ENRESSH. He was also very supportive of young researchers and their career development. Paul's work and memories about him will stay with us for a long time. This Round Table discussion is probably one of the last contributions made by Paul.

Tim C. E. Engels (University of Antwerp) is head of research affairs & innovation at the University of Antwerp, Belgium, and head of ECOOM-Antwerp, the Antwerp branch of the Flemish Centre for R&D Monitoring (ECOOM). In 2008 he started the preparations for the Flemish Academic Bibliographic Database for the Social Science and Humanities (VABB-SHW), the tenth version of which will be released by ECOOM-Antwerp in 2020. Within the European Network for Research Evaluation in the Social Sciences and the Humanities, he has been

leading the work group on databases and the uses of data for understanding SSH research. He holds a PhD in developmental psychology.

Ioana Galleron (Université Sorbonne Nouvelle) is a professor of French literature and digital humanities, with a specific interest in the theatre and in 17th-18th century works. From 2009 to 2012 she has been a pro-vice-chancellor for human resources and finances at the university of South Brittany. She is actually a member of the council supervising study curricula and student life at Sorbonne-Nouvelle University, as well as the head of the disciplinary commission. She is particularly interested in the diversity of research outputs of SSH scholars, and in the evaluation of new and interdisciplinary research in the SSH. In this respect, she is preparing a special issue of the journal *Word and Text*, dedicated to the “New Humanities” (t. b. p. October 2020).

Emanuel Kulczycki (Adam Mickiewicz University in Poznań) is the head of Scholarly Communication Research Group and the director of the Doctoral School in the Humanities at the Adam Mickiewicz University in Poznań, where he is appointed as an Associate Professor. Moreover, Emanuel is a policy advisor for the Ministry of Science and Higher Education in Poland. He got his PhD (2011) and habilitation (2016) in philosophy.

Michael Ochsner (ETH Zurich & FORS Lausanne) is a sociologist working at ETH Zurich and FORS Lausanne, Switzerland. His research focuses, on the one hand, on research quality in the SSH, comparison of national research evaluation systems, and methods of peer review. On the other hand, he is a survey methodologist and works on comparative welfare state research and on labor division in couples.

Gunnar Sivertsen (Nordic Institute for Studies in Innovation, Research and Education, Oslo) is Research Professor and Head of Bibliometric Research. Sivertsen established bibliometrics as a field of research in Norway in 1988 and has since then frequently contributed to the core conferences and journals of the field. He has impacted not only the academic field of bibliometrics and research evaluation, but also advised the development of the research evaluation and funding systems in several countries (e.g. BE, CN, CZ, DK, FI, IS, GB, NO, PO, PT, SE) in direct interaction with governments, institutions, and other stakeholders outside the academic community. He holds a doctoral degree in 18th Century Scandinavian Literature.

Jolanta Šinkūnienė (Vilnius University) is a linguist and Associate Professor at Vilnius University. She is a member of the Committee of Humanities and Social Sciences at the Research Council of Lithuania. Her research interests focus on disciplinary cultures within SSH, academic rhetoric, research publication practices, evaluation of research, career development in academia, and academic identity aspects.

Geoffrey Williams (Université de Bretagne Sud & Université Grenoble-Alpes) is co-founder and President of the EvalHum Initiative. He is a former Vice President for International Relations at the Université de Bretagne-Sud, France (UBS), and has a particular interest in research evaluation protocols and their effects on the Social Sciences and Humanities. He carries out evaluations for numerous national agencies and serves on numerous evaluation panels in the humanities and open science. A digital humanist, corpus linguist, and lexicographer, he is a former president of the European Association for Lexicography – EURALEX. He is currently director of the Department for Document Management in UBS, and a member of the Digital Humanities group of the Litt&Arts research unit of the Université Grenoble Alpes – CNRS.

Jolanta: You are all founding members of the ENRESSH (*European Network for Research Evaluation in Social Sciences and Humanities*) COST Action 15137, which united over 140 researchers from 37 countries and a number of disciplines. How did it all start and what was the core rationale behind this endeavor?

Ioana: SSH evaluation in France is extremely complicated, for a series of reasons: lack of institutional culture in evaluation of these fields, a specific interplay between higher and secondary evaluation, with the selection process for the second being taken into account in the recruitment for the former, lack of instruments such as databases and negotiated criteria, etc. Both me and my colleague discovered all this when we were pro-vice-chancellors of our institution, in charge of quality processes, respectively of international development. We asked ourselves what can be done, and we started to look at other experiences, in European countries. This is how we met the founding members of the Action, at various conferences, or through personal contact after reading some of their papers. A first conference organized in Rennes in 2012 increased the network and started to put to the fore several shared questions and lines of inquiry. We have gradually realized that SSH evaluation is a problem not only in France, but in Europe and in the entire world.

From there on, we started to look at an instrument allowing us not only to cooperate on the research side, but also to have some organizational impact in our respective institutions and higher education systems. COST actions appeared suited in both respects. It was not an easy application, and it took us three goes to be successful, but the continuous growth of the network during these three years and the exchanges with the members during the preparation phases confirmed us in thinking that such an initiative was necessary and timely.

Jolanta: Geoffrey, is this the way you remember it?

Geoffrey: ENRESSH grew out of EvalHum, so to understand the former it is necessary to understand the latter.

Things started off in 2012, a period when ranking and worries over the place of the SSH in the upcoming H2020 was very much on people's minds. At an event in Brussels, I met Tim Carel Stolker, now Rector of Leiden University, and very importantly Alan Palmer of the British Academy. We spoke together for the simple reason that their panels were completely overlooking the SSH. Something had to be done, but what? The answer came from Alan Palmer who offered Ioana and me the opportunity to run a workshop at the Academy in London. Prestigious location, and a budget not far from zero from our own research funds, but we went ahead and this is where you will find all the founding members of EvalHum and ENRESSH. The fact that people came, the fact that a small group of us who had never met before went off to a pub afterwards to plan just shows that evaluation was a real issue and there were people willing to give their time and enthusiasm to tackle it. It is worth bearing in mind that now with 140 researchers from 37 countries that same sense of enthusiasm is still there, and this is why ENRESSH has been so successful.

Jolanta: Interesting! What were the issues that brought you together?

Geoffrey: They were three: the growth of ranking and its effects on competitive HE, the place of the SSH in H2020, which for the EC at the time was nowhere, and research evaluation with its effects on the SSH.

The three are obviously linked. The other two were different, but often mixed in people's minds so that the second was poisoning the evaluation debate. Ranking was a major issue in 2012 with the rival systems influencing heavily universities and with the EC developing its U-Multirank system. Ranking is still there, but

has lost a great deal of weight on the ground floor as universities realized that their home market and traditions were their primary market, and that the big is beautiful is not necessarily efficient. The plea for internationalism has also changed as immigration restrictions get put into place by the same politicians who called for an open market in Higher Education! As it stands, we did not get directly involved, except through the issue of research evaluation.

H2020 was another issue, and was the reason behind the British Academy meeting, and brought the same people to Vilnius (to the Horizons for Social Sciences and Humanities Lithuanian EU Presidency's Conference, September 23-24, 2013) as who came to London. The net was spreading. This too was essentially a political issue, and very much an ongoing one. What became EvalHum was active, but also active in what is now EASSH and so we have left campaigning to them.

What was becoming clear is that the big issue was research evaluation procedures. How to make them better, how to make them acceptable. The confusion with ranking and the barely hidden political motives have greatly confused the issue and led to a strong anti-evaluation movement, which happily forgets that they evaluate all the time. We knew that evaluation was necessary, but that it had to be fair, and that consequently rather than campaigning on political issues we need to step back and get an overall view of research evaluations across Europe, and beyond. This required creating a community and obtaining funding.

Jolanta: Did you try to apply to COST for the funding straight away?

Geoffrey: Our first attempt was to get funding via the European Science Foundation. We failed after an evaluation procedure that showed that transparency was clearly needed. At this time, the ESF was stopping funding so we could not try again; instead we organized our own Extra Special Forum in Rennes to which everyone came paying their own costs. This was a small event, which was attended by the head of the French evaluation agency, called AERES at the time, Didier Houssin. The following year, we organized the first of the RESSH conferences, and EvalHum become officially an association and such a success. This still did not give us funding, so encouraged by the late Philippe Keraudren of DG Research and Innovation, very much the friend of the SSH in Brussels, we tried H2020 calls, and failed each time at the last hurdle. This was beginning to show us that evaluation procedures needed a close looking at everywhere, and that also there was an institutional distrust in anyone looking at an SSH evaluation.

H2020 was about research funding, but we also needed networking and so we attempted COST, and failed, twice. The file was getting better every time, but

insider knowledge told us that the SSH were not a priority and that research evaluation was not a welcome subject. We were lobbying hard and greatly benefitted from the assistance of our French COST NCP, Emmanuel Pasco-Viel. On our third attempt, I asked Ioana to lead the project on the basis that maybe the gender factor would work. It did. We knew our file was brilliant, we know the subject is important, but we know that the internal battle was hard to get us through. The news came one Friday evening when Ioana and I were heading back to the airport after a conference in Lyon. The first thing we did after check in was to order a glass of champagne. The rest is history.

Or rather it isn't. History is the past and ENRESSH is very much the present and the future. Our kick-off meeting in Brussels brought together the initial team, our members from the first countries involved, and set off a process that COST can be proud of as a very successful action under the chair of Ioana and then Emanuel. The action is over but the work continues, as does Eval Hum.

Jolanta: You are right, ENRESSH is the present and the future, which builds on a successful past. Now that the Action has come to an end, how would the leaders of the key work groups reflect on the results that have been achieved? Here I am addressing Michael, Paul, and Tim. Have the results provided answers to the questions you had at the beginning of the Action, or on the contrary have they raised additional questions?

Michael: I have just finished the report for work group 1 on conceptual frameworks and also the overall report for the Action, which is a good opportunity to reflect on what were the initial ideas back in 2012 when we started discussing what would later become ENRESSH, how these ideas evolved until the start of the Action, what were the feelings during the Action when you coordinate a large group of scholars, and what are the final reflections on the results ENRESSH has achieved. It is very interesting and also enlightening to see the different stages of your own perceptions. It starts with concrete ideas on research projects, some dreams on interaction with policy makers. It went on, in the stages of different proposals, by enlarging the network, which was already a success in reaching out. When the action started, I had some goals I wanted to achieve in mind, some publications mainly. I thought that the collaborations will increase my publication output, but in fact, the opposite happened. I used most of my time for emails, reports, and updating and merging files of different collaborators etc. I also struggled in motivating people to work – without funding for research – on the topics we wanted to find solutions for. We discussed hard among the different stakeholders

within the Action about sense and nonsense of certain terms and concepts or even research projects. And in the end, when writing the report, I realized that all those discussions and detours were actually more valuable and relevant than what my own ideas were in the beginning; they were the real results.

Jolanta: So those four years did not turn out to be exactly as you expected. What were the most unexpected aspects of the work in the Action that you could not possibly predict?

Michael: The most unexpected result during the Action was that I made so many new friends. Indeed, our network was not only about work, I spent a lot of leisure time with fellow ENRESSHers. There was not much competition but an incredible productive, constructive and friendly collaboration, not just cooperation. As collaboration takes much more time than cooperation (you need to agree on every step as everyone is taking responsibility for the whole, not only for parts of it), I worried a lot about outcomes during the Action. Especially, the negotiations to arrive at our first Guidelines for SSH Research Evaluation in the first year were frustrating as there were so many different opinions and interests. I was quite worried because my group had to produce a policy brief that went even further: to provide “better adapted criteria for research evaluation in the SSH”. So, I started early to remind our members that this task is in the making. In the last year, I put it as a main topic on the agenda and was worried about how it will go and I was amazed how we had a productive brainstorming. It was a turning point for me, as I started to understand what we will achieve. I drafted a first version of the policy brief, putting much of what I thought were my own priorities into it because I expected that the discussion would water it down. However, during the consultation in the task force I only received comments regarding details, mostly the opposite of watering down but rather making the messages more explicit. The consultation in the bigger group also was only making it more precise and it became clear to me what ENRESSH actually achieved: we had four years to work together and have developed a common, shared understanding of what needs to be done; we came from different corners and through mutual learning, we all developed into the same direction. We had so many discussions across different stakeholders involved in ENRESSH that we were able to formulate recommendations that work for all of us; there was no negotiation, we actually agreed anyway, there were not “my own ideas” and “the other one’s ideas”, we discussed so much that it all became “our ideas”. This is an incredible achievement I would never have thought possible to be reached. This example summarizes what ENRESSH is to me: we have done much

research on this and that topic, but we all worked towards the same goal, which is finding evaluation procedures respecting, supporting, improving and, importantly, valorizing SSH research. While often in research, one gets lost in details and cannot tell the forest from the trees, in ENRESSH it is the opposite. All projects were part of something bigger. The small projects were interesting but only means to an end. ENRESSH is a very diverse network regarding disciplines, stakeholders, countries, and European regions. But we have managed to set a common agenda. The COST Action is over, but I am sure we have just started.

Jolanta: Indeed, thinking in the same direction seems to be one of the greatest achievements of the Action for its participants! Work group 1 was also one of the largest groups in the Action, so it is amazing that people from so many different backgrounds came to thinking in a similar way about SSH evaluation. You also mentioned valorization and I immediately thought about the societal impact and relevance of SSH research which was another major focus of the Action, tackled by Paul and his work group. What are your impressions, Paul?

Paul: The interesting thing for us as a COST Action from my perspective is that we have been attempting to integrate and engage with existing research activities. The question of how does social sciences and humanities research create impact in society is a question that has only become more urgent during the programming period. We have seen a range of research funders invest in their own research activities, and at the level of the European Commission, there have been a series of research projects within the “Science with and for Society” programme. It is extremely gratifying that these projects have encountered what we took as our starting point, namely that this issue of SSH research impact is extremely complex. But by not taking that complexity seriously, and assuming it can be simplified in line with political and policy imperatives, those other research activities have not come any further in producing answers.

It can be easy to become enthusiastic about the opportunities offered by new measurement phenomena like Altmetrics for their alluring simplicity and promise that they are able to capture in a simple number some characteristic related to the “goodness of the research.” But that simplicity at the same time is a weakness in that these alternative impact metrics are easily gamed and weakly linked to the underlying knowledge process. They rapidly lose their value once they start to incorporate more of the dimensions that underpin societal impact, as does the proprietary Plum metric. And indeed as the work of Jack has shown, any kind of

societal impact tracing approach can be time-consuming and generates a very hermeneutic version of impact that is totally unsuited for mechanical and unthinking distributive evaluation approaches.

By foregrounding that complexity and the complexity of the answers that the research produces we have been able to progress and not find ourselves bogged down in what are essentially infeasible and irreconcilable demands from policy makers. Where we have been able to make progress has been the more general question of why policy-makers seem so wedded to simplistic models of research impact evaluation, and ultimately to understanding how we might be able to better address that in the future. The future of impact evaluation has to be more context-sensitive and more formative if it is to have any kind of future at all, because otherwise you simply run the risk of rewarding lucky people, people who can tell a convincing story about their impact, and not those that have worked effectively with societal partners to create knowledge with the potential to become useful later.

Jolanta: The work on impact done within the Action would hopefully find its way into the future European research and innovation programs and frameworks, which seem to become increasingly more impact-oriented than the previous ones. Last but not least, the third work group led by Tim delved into the world of bibliographic databases and uses of data for understanding SSH research. What's your perception of those four years of the Action, Tim?

Tim: The Action has greatly advanced the state of the art. Prior to ENRESSH, we knew about national bibliographic databases in only a handful of European countries. Soon after the Action started, Linda Sile had identified over twenty. We went on to compare publication patterns across countries in a series of papers, among other things showing the large differences in publication patterns also within disciplines (Kulczycki et al, 2018), the remaining prominence of book publications (Engels et al, 2018) and the importance of the use of multiple languages in SSH scholarly publishing (Kulczycki et al, 2020). In parallel, we worked on a proof of concept of a European Scholarly Publication Infrastructure (Puuska et al, 2018) and published a good practice manual for national bibliographic databases (Sile et al, 2019). Many new questions and issues have arisen too, yet thanks to this COST project a strong European community has now been established.

Jolanta: There seem to be indeed many significant achievements of the Action in various respects. However, what has come up in Tim's reply is an

important idea of differences within the disciplines of SSH. Indeed, the fields of social sciences and humanities are generally very diverse with epistemological, theoretical, and methodological traditions ranging across different disciplines and across different cultures. How are we to reconcile these differences in designing adequate policies for research evaluation and valorization?

Emanuel: Designing policies for research evaluation and valorization is a complex task which should be started by addressing the following question: What is the research evaluation for? Experience derived from numerous discussions with ENRESSH members has taught me that research evaluation is too often done just for legitimization of policy decisions like funding distribution and categorizations of research institutions into good and poor performers. Research evaluation should, first of all, serve both those who evaluate (by supporting realization of their goals) as well as those who are evaluated (by providing useful feedback). It means that designers of good evaluation procedures have to consider how evaluation methods can be suited to the evaluated object and how the results of evaluation and the feedback would be useful for those who are evaluated. Policy designers should also know that by designing rules they constitute objects of evaluation and this has relevant practical implications. One can clearly see it from the example of ‘research excellence’, a concept which is difficult to understand and define without clear characteristics and variables by which one can measure it. For instance, one institution conducting evaluation can define it through a combination of indicators (e.g., field-weighted citation counts), whereas another through a more qualitative way expressed by a short description of one of the most important achievements within the last five years. From the perspective of those who are evaluated, research excellence—as a criterion of evaluation—is real only when clear criteria suited to them are implemented. Otherwise, research excellence is a black box to which no one (even its designers!) has access. Such a type of evaluation is useless.

Various ENRESSH studies show us that policies should be suited not only to the relevant fields of science but should also take the geopolitical dimension into account. For instance, internationalization policies in non-English speaking countries should be designed with a consideration of current publication patterns observed in countries having different histories. Countries like Finland and Norway are non-English speaking countries; however, their internationalization has a different form than the internationalization of Central and Eastern European countries.

Jolanta: So not only disciplinary but geopolitical aspects as well should be taken into account when designing evaluation policy?

Emanuel: Research evaluation requires many different decisions (e.g., what kind of indicators will be used). It is up to them whether it will be a good procedure or not. Thus, in designing policy, it would be good to base it on a few clear principles that take into account both disciplinary and geopolitical differences. Firstly, research evaluation should not be (only) publication-oriented because researchers and their activities cannot be reduced to publications. Good research evaluation covers various dimensions (including the societal impact of research) and values (e.g., cooperation in very competitive academia). Secondly, all publication types matter because different disciplines have different publication practices. A good research assessment allows researchers to report various publications types (monographs, book chapters, proceedings, journal articles) and to use the most appropriate databases for the field. Thirdly, good evaluation procedures should not be based on a single indicator. More (but not too many!) indicators should be used to increase the uncertainty of evaluators' decisions which in the end would increase the quality of evaluation. Fourthly, designers of evaluation policies should remember that science has many international languages and multilingualism keeps locally relevant research alive. Thus, good research assessment supports the dissemination of research results in various languages. Fifthly, by evaluating researchers, we show them what good evaluation should look like and in this way we teach early career researchers how to be a peer and an evaluator. All of us should remember that no one is born an evaluator. Finally, research evaluation legitimizes publication channels and creates incentives. Thus, by showing what publication channels (e.g. top-tier journals, open access publications, etc.) are counted in evaluation, evaluation creates very powerful incentives.

Jolanta: You have just mentioned publication channels, but there are so many of them. Can registers and repositories be of any help in facilitating research evaluation? A case in point could be, for example, Academic Book Publishers (ABP). How can this tool assist policy makers in designing guidelines for evaluation in SSH? We have to address this question to Gunnar, who is one of the founders of this register.

Gunnar: Academic Book Publishers (ABP), a global and multilingual register, is being built as an interactive and dynamic register of scholarly book publishers who support the research quality standards of the SSH in their peer review and

publishing practices. With this initiative we aim to defend and improve the quality standards of scholarly book publishing, and make these standards reflected in proper research evaluation procedures.

ABP shares objectives with other initiatives such as DOAB, OPERAS and *Think. Check. Submit.* We supplement these initiatives with:

- A broad coverage of scholarly book publishers, including the large international imprints as well as the smallest at the local level (more than 7000 imprints from more than 100 countries).
- Hence, a broad coverage of languages represented through these publishers.
- A bottom-up approach with contacts and sources in all countries, building the register from national and complete data sources to an international master list, which guarantees the diversity of book publishers represented.
- Quality standards or validation systems for inclusion.
- A working team in direct contact with both the academic book publishing industry, the research performing institutions, the research evaluation and funding agencies, and governments.

Our idea is that the information in the register will come from several independent sources: from the publishers, from national bibliographic databases and legal deposit libraries where publications from research institutions are recorded, and from the scholarly community itself by feedback either given directly from the authors or through Current Research Information Systems.

See also: <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/05/bibliodiversity-what-it-is-and-why-it-is-essential-to-creating-situated-knowledge/>

Jolanta: I see, so in a sense it could be treated as a reliable collective database of quality publications that can perhaps be consulted not only by those in charge of research evaluation, but also by scholars who are looking for a potential publisher. Indeed, it is very useful for evaluation in many respects. What recommendations in general would you give to policy makers and research managers concerning criteria and methods for assessing SSH research?

Gunnar: My answer to this question starts with describing a problem and ends with suggesting its solution:

The problem is that certain commercial journal indexing services with relatively poor coverage of the scholarly literature in the SSH are regarded as ‘top standard’ or even as sufficient information sources in evaluations. The presence of publications in Scopus or WoS has increasingly become a criterion in evaluations of research in the SSH. Some countries have even installed protocols for research evaluation

or performance-based funding models where publications that are indexed by the commercial databases are treated separately in indicators of “internationalization” and “research quality.” In other countries, there is a general belief that research quality can be promoted in the SSH by expecting more publications in the limited number of international journals that have been selected for indexing. Consequently, for several years already, Elsevier and Clarivate Analytics have experienced a pressure from researchers in the SSH to have more journals indexed. Both providers have responded by increasing the coverage of journals and book series, and, recently, even of books in the SSH. However, the coverage of the scholarly publication output in the SSH is still limited, as has been demonstrated in several studies published by the ENRESSH network.

The shortage is mainly due to the more heterogeneous scholarly publication patterns in the SSH where publishing in international journals is supplemented by book publishing and the use of journals in the native languages. Just as with the abuse of Journal Impact Factors in research assessment of individual performance in the natural sciences and engineering and in the health sciences, the “coverage criterion” in the SSH represents an artifact which is external to and beyond the control of the scholarly norms and standards that it is sought to represent. It creates unnecessary tensions between fields in the SSH with different degrees of coverage in the databases. It also creates debates about what will happen to the use of books and native languages in the SSH. In these debates, the general development toward publishing in journals covered by Scopus or WoS is often perceived as “inevitable” and driven by new evaluation regimes, not by internal scholarly standards.

I propose an understanding of the processes of internationalization in the SSH which is independent of the “coverage criterion” and instead related to concepts of field-specific research excellence and societal relevance in the SSH. In a historical perspective, it is easy to demonstrate that the SSH are not originally “national” in their publishing practices. They started by being international within an academic elite. In Europe, Latin was the first of several international languages that have been used during several centuries. The “nationalization” of the SSH is closely connected to the democratization of education and cultural and social life in the 20th century. Today, the quality and relevance of research in the SSH are checked not only by peers but also directly by society. Internationalization is important for research quality and for specialization on new themes. Interaction with society is just as important for realizing the ultimate aims of knowledge creation. Taking both purposes into consideration at the same time, there is no reason to apply a general hierarchy of languages or publication types in the assessment of research in the SSH. All the communication purposes in all different areas of research, and all the languages and publication types needed to fulfill these purposes, should be

considered in a holistic manner without exclusions or priorities whenever research in the SSH is evaluated.

Jolanta: Thank you for an extensive answer that takes into account various perspectives of evaluation. What advice can we give to researchers in SSH who are frequently lost in the changing or unclear rules of the game? Are there any ways in which they can contribute to the quest for the best path to research evaluation in SSH?

Paul: The basis of any kind of research evaluation is to help guide scientific decision-making processes. Research evaluation is not about economic efficiency, but rather is an attempt to help improve a decision-making process. That process involves scientific communities deciding what are better or worse activities, to channel resources to the “best” activities which best contribute to advancing the scientific state-of-the-art. What good evaluation therefore requires is a common shared understanding of what constitutes good scientific practice, and that is the role played by peer review in research evaluation.

What ENRESSH has been concerned with is dealing with the problem that policy-makers have been quick to impose systems that reflect a common shared understanding of a fraction of the scientific community – think of research evaluations that are based on the common practice in natural science fields of publishing many shorter papers each with simple messages written by multiple authors. If that is your view of what is good research, then there are swathes of the social sciences and humanities for whom that evaluation has no relevance. That evaluation approach cannot meaningfully work as a steering technology. If for example as happened in the UK with the Research Evaluation Exercise, then a discipline like management sciences can find itself simply churning out simplistic garbage to satisfy this external requirement, undermining the relevance of the findings to the field. So what researchers need to do in the SSH to ensure good impact evaluation is two things.

Firstly, they have to develop their own sense of what constitutes good societal impact within their field, bearing in mind that it cannot be dependent on the luck of knowing Prime Ministers, but has to be achievable by many researchers. So there needs to be an understanding and acceptance of the ways in which impact creation activities are built into the tasks that SSH researchers carry out in their research, and of the fact that these are extremely diverse, strongly related to particular kinds of research, and not necessary for all researchers to undertake. This is a rather challenging demand in itself, and needs to begin from the Ph.D. phase,

providing doctoral researchers with the opportunity to see the everyday ways that research is embedded in society, and to understand how to actively manage those activities, even where they are not themselves immediately active. And when SSH researchers become more senior and move into decision-making positions, they need to be willing to recognize those everyday impact generation activities as being valuable, to encourage SSH to collectively evolve to become more conscious of their impact-creating activities.

The second issue is something that is not completely in the hands of the SSH researchers themselves, but it is clear that there is a need to change the way that SSH impact creation is viewed by external parties. All too often, everyday engagement is seen by external actors as something that is not intrinsic to good social sciences and humanities research practice, but rather than is something undertaken by academics who are not good enough to do the proper business of research. This discourse regularly crops up where evaluation techniques are used to make allocative choices about research funding, about promotions, appointments, and fellowships. It can be relatively straightforward for SSH researchers to make the claim that their monographs are as valuable as short articles in international journal. But claiming that writing a museum catalogue or working with a community group is equivalent to working with an innovative business remains extremely difficult, both to other scientists and to policy-makers.

Michael: The most important message to SSH researchers is: do not copy the STEM disciplines and do not trust any of the commercial suppliers of data and evaluation products. Their methods do not at all work for SSH research. Furthermore, confronted with our results for the SSH, STEM scholars usually agree with our criticism of the simplistic procedures and tell us that it applies to STEM research just as well. It is just that STEM researchers are used to quantify and thus the commonly used approaches are closer to their methods. They are not used to the fact that their object of study is reacting to the study results, which is of course the case for research evaluation. It is the very reason why we do evaluation, in fact. If nobody would react upon an evaluation, it would be a useless tool for policy making.

A second message is that in the SSH, books continue to play a role. Do not stop reading and publishing books and book chapters simply because the STEM disciplines do not value them and because they do not have Impact Factors. SSH disciplines have a different way of producing knowledge and journal articles can only present partial results of a research project. Indeed, really good research addresses complex issues and therefore needs a book to be adequately covered. Some of my younger colleagues told me to stop reading and publishing books as this is old-fashioned, takes too much time, and won't help in getting a professorship.

Book publishing is not old-fashioned. Not reading books because an article is read faster is just another way of saying that you compromise on research quality.

A third message is to always be cognizant of our own limits of expertise, especially in interdisciplinary contexts but also when discussing with colleagues from the same field following different approaches or paradigms. Other disciplines or traditions have different approaches to research that might be different but no less valuable for their purpose, and evaluation procedures need to account for those approaches. Also, research evaluation is a discipline on its own and SSH scholars should not just apply indicators mechanically but leave this to the experts. They should forget the h-index, not use it and tell everyone not to use it. They should reflect on the data quality behind science indicators and try to reflect it in relation to their own expertise: if the data they use for their object were of similar quality and indicators of similar precision or validity, would they use it? Especially, the Altmetrics Doughnut or Research Gate Scores should not be used.

A fourth message, out of my own frustrations with my co-authors and as an associate editor of a journal, regards Impact Factors. Never use Impact Factors (IF) for the selection of the journal where you want to publish. Journal selection must follow the content of the article and the audience it is written for. It does not make sense to submit an article on attitudes toward welfare state in a journal for computational social science methods, also when the article is really good and it's worth to publish it in a "high-quality" journal in the social sciences. If there is no computational social science in it, it should never be submitted to this journal even though it has a high IF in the category "social sciences." I even doubt that it makes sense to use the IF in journal selection within the narrow range of "valid" journals (in this case welfare policy), simply because the IF does not tell anything. Rather, publish the article in the journals you read when you want to learn something about the topic. Otherwise, you create only work for editors who have to desk-reject articles that do not fit at all the journal's scope.

A fifth message concerns the hype around societal impact. I was obliged by my institution to follow a short course in "societal impact creation." The teacher told our faculty that before starting a research project, we should call the ministry and ask them what they want as a result. This will help to reach societal impact. Each doctoral student should call the ministry and tell them their results when they're finished. Imagine the busy telephone line of the poor minister who will hate universities within a few weeks, the only societal impact would be that the ministry would not be effective anymore. We have to stay humble. Not every single research project must have a societal impact in the sense of actual impact on policy or a commercial product. Often, for example, it is very valuable to know that doing something is of no use. Thus, having no impact is also impactful. Furthermore, it

is not a project or a scholar that/who needs to have an impact. It is the discipline as a whole. Maybe a big, multi-year, multi-institutional project can be expected to have a visible impact. But generally, knowledge creation is a collaborative process. It makes no sense to think that a single person can make a difference. As a consequence, prizes to individuals are useless, even counterproductive. They create a wrong idea of what research is. We should get away from the cult of the genius. There is only one stable genius in the world anyway. And, honestly, do you want to be like him?

Emanuel: In my opinion, most researchers perceive evaluation only as a bureaucratic burden which is a constant reporting combined with filling out various forms. Thus, if they can contribute to discussions on the type of research evaluation in their institutions, regions, or countries by providing information on what is not working and on what is good in current procedures, it would be a very useful starting point. Some years of working as a policy advisor have taught me that only suggestions which are written down in the form of some text, clearly communicated and delivered to the policy makers, are taken into consideration. Thus, my advice would be: just let the persons responsible for evaluation know what is not working and how, in your opinion, it could be improved.

Ioana: The difficulty with research evaluation is to try and judge (a piece of work, a project, a career, etc.) in a less idiosyncratic way. There is no such thing as perfect objectivity when it comes to research evaluation, but this does not mean that one cannot strive towards it, or rather towards an informed judgment. Bad things happen in research evaluation when a person or a small group become the reference point, whose habits and way of thinking constitute the absolute “must” to the detriment of others. So, in my opinion, researchers could contribute to improving research evaluation in the SSH by opening to the others – from other fields, with other approaches, other practices. Seeing how research is done elsewhere is not only inspirational but can help one put things into a bigger picture when it comes to the expectations involved in the evaluation exercise. The amount of different work experiences and research settings should be even a criterion for picking up evaluators, and in this respect I daresay that some early career researchers are better equipped than some seniors who have spent all their life in the same institution.

Tim: Be open-minded. Too often, the SSH are positioning and being positioned as different, whereas this is far less often the case than is commonly thought. Interdisciplinary work and other ways of interacting with scholars from fields far

away from one's own are the best paths to understanding how research, and hence research evaluation, works in different fields. One can make a major contribution by broadening the horizon.

Jolanta: Now that the ENRESSH COST Action has come to an end, how do you see the future of research evaluation in SSH and the future of the network that has been created within the ENRESSH?

Tim: The current pandemic shows again how necessary an interdisciplinary approach is. In these interdisciplinary collaborations, SSH plays a major role. Where this is not yet the case, it will become a reality sooner or later. Therefore, I see a lot of future potential for our network, as it reaches well beyond the SSH *stricto sensu*.

Emanuel: As ENRESSH, during the four years, we have achieved much more than I could expect. ENRESSH as a network of researchers, policy makers, and friends will be sustained in various forms. In these upcoming years, the expertise and tools produced by ENRESSH might be very useful. Governments and societies have been noticing the importance of research to fight global diseases like COVID-19. However, it is time also to constantly highlight that finding vaccines is one of the tasks for science these days. The other task is to serve society and to provide the understanding of the current transformations of social relations and economy. In this, the role of SSH cannot be overestimated. Therefore, the future of research evaluation should highlight the role of SSH in science and higher education landscapes.

Ioana: ENRESSH finishes in a very uncertain time, with many plans being blown by the wind of epidemics. The natural extensions of our activities are the EvalHum association and the RESSH, but we don't know for the moment when the next one will take place. For the moment, we need to stick together via initiatives like this: virtual gatherings, remote collaboration, on-line publications. I also expect that smaller forms of cooperation will take place in the following months and years, between pairs of researchers in the network. Also, ENRESSH should maintain and even increase its presence in EASSH, so that it remains a voice listened to by the EC.

Michael: I think that the future of research evaluation is bottom-up, linked to research practices in the respective fields. I hope that validity becomes more

important when it comes to measurement and I wish that conceptual thinking finally finds its way into research evaluation: what do we want to achieve with the evaluation, how can we achieve it and then, how can we evaluate whether evaluation has achieved its goals? The measurement-approach has failed using bibliometrics, it has failed for altmetrics as well. Now they try it with societal impact and will fail. I am confident that sooner rather than later, science managers will understand that measurement can only work when one knows what one wants to measure. Measures do not help to find out what one wants to measure. More indicators do not give you more information, rather they blur it. Information that is not correctly contextualized is not information but disinformation. Think of Covid. The number of positive cases is disinformation if you do not know the number of tests made. If you add the number of cases per day, you do not know better if you do not know whether the testing policy changed, and so on.

There is no way around for leaders to develop and follow a vision, to convince others to follow this vision, and then to take responsibility for the outcome. There is no “following the numbers” and outsource responsibility to numbers. Decision making means that decisions have to be taken. There is not “one truth” represented by the numbers. SSH scholars know that and have the duty to explain it to science policy makers: all is dependent on the context. The same scientific result will be interpreted differently across time and cultures, we see it in the Covid-crisis (see, e.g., the very different research-based policy decisions on wearing masks. It varies across cultures – in Asia it is clear that it protects spreading it; in some European countries it is clear that the mask does not protect you getting it – but also over time – in Switzerland and France masks were said to be scientifically proved useless and counter-productive when there was a lack of stock and two months later they wanted to introduce the population to wearing masks, with more or less success, unsurprisingly).

ENRESSH is here to exactly do that: contextualise. We are a large network that has acquired a lot of knowledge about research evaluation but also about knowledge production across Europe. The network includes members of many different stakeholders and we have managed to find a common language and a common argumentation. We all will carry on with working on improving evaluation procedures and will fight for SSH budgets. But we will certainly also continue to do research together as we have produced a lot of data that still needs to be exploited. Some will be parting ways; some new people will join. We will stay connected via our international association and our bi-annual conference.

Paul: There are a lot of “zombie ideas” in impact research evaluation that simply can’t be killed, and even after generating deep understanding and strong arguments

for context-specific research evaluation, these zombies keep resurrecting and demanding simple comparable metrics often based on a simplistic understanding of more technical disciplines. The key issue is that the way that research takes place is changing continuously, opportunities to work with societal partners, to co-create and co-determine knowledge is evolving continuously. Sometimes that is evident when a big trend gets a name attached to it, such as digital humanities. This created a whole new approach for academics to work together with societal partners to create impact from their ongoing research activities.

These changes in research practice that affect the way that researchers are integrated into society in their daily working practices are happening across all kinds of research, not just when this is explicitly identified, and so the goalposts for impact evaluation are themselves continually shifting. This provides a challenge for ENRESSH into the future, but also a strong case for its continuing relevance. The message of the last decade is that demanding that SSH research is recognized and included in research impact evaluation is an ongoing urgency, to ensure that science policy debates do not become excessively narrow and instrumental, not just to benefit the recognition that SSH gets, but that all sciences get for their profoundly enriching impacts upon contemporary society.

Jolanta: Thank you very much for this interesting and insightful Round Table discussion!

Moderator of the Round Table Discussion: Jolanta Šinkūnienė

ACKNOWLEDGEMENTS

This paper is based upon work from COST Action CA 15137 'European Network for Research Evaluation in the SSH (ENRESSH)', supported by COST (European Cooperation in Science and Technology).