SUMMARY. Every researcher independently of the field she represents takes part in the evaluation processes. Evaluation is crucial for the Social Sciences and Humanities (SSH), as it might either improve or worsen their academic reputation. In the article we discuss a COST Action ENRESSH (CA-15137), which sought to analyze the SSH evaluation systems, practices, and criteria, as well as their impact on the SSH societal status. The aim of the article is twofold: firstly, it attempts to overview ENRESSH’s most important outcomes and major findings; secondly, it seeks to reflect and evaluate the action’s impact on a better understanding of the evaluation phenomenon in SSH.

KEYWORDS: ENRESSH, social sciences, the humanities, research evaluation, diversity, national languages, societal impact.

In April 2020 one COST Action mostly known by its acronym ENRESSH has officially come to an end. ENRESSH stands for European Network for Research Evaluation in the Social Sciences and the Humanities (CA-15137). According to the final report and its official approval, it did not end in tears. It was rather a big success, at least in the opinion of its participants. We scrutinize this claim by linking the Action’s goals and results available from the ENRESSH website\(^1\) and relate this to the current practice of research evaluation. But beware, the authors of this article are themselves committed ENRESSHers and by no means neutral.

SHORT DOSSIER ON ENRESSH

In the Memorandum of Understanding\(^2\) it is stated that ENRESSH “aims to propose clear best practices in the field of SSH research evaluation.” Its main aims are:

---

\(^1\) See: http://enressh.eu/ [accessed on May 3, 2020].
\(^2\) See: https://enressh.eu/about/ [accessed on May 2, 2020].
• to improve evaluation procedures in order to take into account the diversity and the wealth of SSH research;
• to make a robust case for the ways in which the SSH add value to the society;
• to help SSH scholars better appropriate their research agenda and overcome fragmentation.”

The action started in April, 2016. It was initiated by 20 scholars from 16 countries.

The final number of the ENRESSH countries: 36 (13 high R&I performing countries, 15 low R&I performing or widening, or Inclusiveness Target Countries, and 8 unclassified countries). From the list of the big high-performing players only Sweden was absent, which, as rumors went, was caused by certain Swedish inner policy issues. For this reason every visual geographical representation of the ENRESSH findings had an empty spot in the place of Sweden.

Number of participants: 125 (from high performing, from low performing and from unclassified countries) with the gender balance approximately right.

EXPERIENCE IN AND OF ENRESSH

What did we learn from the action? Lots of interesting if not unexpected stuff. Researchers are a communicative folk. An accidental chat over a cup of coffee or a conversation during dinner or afterhours might effortlessly lead to an unexpected collaboration. Or at least to sincere promises to collaborate, which one might regret later. Why? The majority, if not all, of the participants experience chronical lack of time. Everyone is extremely busy with enormous amounts of deadlines constantly threatening their well-being and mental comfort or at least (accusingly) lurking somewhere in the background. But, of course, it is not an obstacle to engage in another promising and interesting academic endeavor, thus increasing the timeline pressure. Despite belonging to different countries, traditions, or disciplines, people share certain common traits, such as, to put it mildly, not fully rational time planning skills or super abilities to finish preparing one’s presentation slides exactly one minute before the presentation itself. An attentive observer could have a glimpse at more exciting matters, such as the teaming-up patterns, conversation etiquette, communication patterns, divisions along qualitative and quantitative lines, expected and unexpected tensions, alliances, rivalries, friendships… But is all this any different from any other COST Action?

If we compare ENRESSH with other COST Actions, on the surface there seem to be not so many differences: participants from various European (and not only European) countries, meetings, steering group, management committee, minutes,
deliverables, outcomes, conferences, publications, meetings with stakeholders, newsletter, SIGs, ECIs, STMs, etc. But looking deeper below the surface there are certain peculiarities in ENRESSH. One of the greatest things in ENRESSH’s topic is its inclusiveness. Research evaluation is by no means limited to certain sciences or disciplines. All the researchers constantly take part in evaluation processes in different roles as the evaluators, the evaluated, the bystanders, and so on. Thus, nearly every scholar (in SSH as well) has her own experience and (strong) opinion on evaluation procedures, types, principles, criteria, need, justification, aims, etc. And this makes them qualified to participate in ENRESSH as both an expert and a learner... Furthermore, there are no geographically unsuitable countries. Or to paraphrase the famous Monty Python song, every approach or every perspective is sacred; OK, let’s make it “valuable” instead of “sacred.”

The first impression before ENRESSH might be the following: research evaluation is research evaluation everywhere. Let’s discuss how it is done and how to improve it, and that’s it. Usually, evaluation is perceived as a structurally simple process (E – C – O – R), during which E (the evaluator) by applying C (assessment criteria) assesses O (outcomes) produced by R (the evaluated researcher). The end of the story?

Not everything is that simple. Even those who expect differences in research evaluation practices (be they geographic, political, disciplinary, or other) are not ready to open the Pandora’s Box of actual research evaluation practice and prefer it to be shut and stay with E – C – O – R without specifying it much further. Of course, this is not similarly applicable for those few who specialize in the field of research evaluation studies. But even they prefer not to open the Box but rather to circumvent it. Yet the Box is out there. Why the Box? Why Pandora’s Box? Nobody knows what exactly is in the Box. Nobody suspects the quantity or amount or value of the contents. And what if the Box contains things that are better kept locked? Or if the information inside is impossible to contain and process?

ENRESSH was proposed by a group of some 20 researchers from 16 countries, it currently brings together more than 125 participants from 36 countries. Why is this fact important? Or what happens when SSH scholars from 36 countries gather in one place during a meeting? Loads of information, that’s what enter the stage – because all the participants could provide possible data from their respective disciplines and countries. Data, typologies, policies, narratives, success stories, sad stories, stories about passion, stories about injustice, interviewees, you name it. The paradise of, and for, comparative studies.

Could ENRESSH be likened to a kind of Eurovision song contest? To some extent it is possible to find similarities between the two: internationalization, fun, factions, different interests and incompatible musical ... sorry, academic tastes;
and in both, some behavioral patterns can be discovered, etc. What ENRESSH obviously lacks, is a competition among countries and regions. It’s common folk knowledge that a significant part of Eurovision voting is based on geopolitical priorities. Sometimes the competition is even nicknamed “Balkanovision.” It is rather obvious from the number of points given by Greece to Cyprus and vice versa. There is no guarantee that researchers from neighboring countries of the region would love their neighbors, i.e., vote for each other or support each other. It is thus possible to imagine potential alliances between representatives of mutually antagonistic countries if they share a common disciplinary background. Moreover, neighboring countries might strongly differ in their idea and system of the SSH research evaluation. One of the shocks of ENRESSH was Belgium. In sports or cultural or pop competitions Belgium is a single indivisible unit. But when it comes to research evaluation there is no Belgium, there are Flanders and Wallonia (and Brussels). And their evaluation differs more than those of Lithuania and Norway.

WELCOME TO THE D-LAND

What did we learn from that? What is the outcome of ENRESSH? Is it possible to express the major idea of ENRESSH with the help of a limited number of keywords? What keyword was heard most often during the Action? The keywords that easily (maybe even too easily) come to mind are: research, evaluation, SSH, typologies, systems, peer review, publications, stakeholders, policy, (societal or social) impact, productive interactions, metrics, altmetrics, open science, open access, learned societies, young scholars, etc. Is it possible among these keywords to extract “the one”? That is a matter of debate. But from our perspective the one (and only) is DIVERSITY. Wherever one looks, it is impossible to avoid the diversity of something: diversity in evaluation systems, linguistic diversity, diversity in publishing patterns, disciplinary diversity, etc. What tells an ENRESSHer apart from others involved in research evaluation, is that diversity is not seen as a problem but as the state of the art in SSH research evaluation.

DIVERSITY IN EVALUATION SYSTEMS AND PROCEDURES

An ENRESSH-based research group tried to dig deeper into national research evaluation systems (Ochsner et al. 2018). The data was collected from 68 participants representing 32 countries. After thorough analysis eight major categories were singled out: “national database” (database covering national research outcomes
in/of a given country); “SSH specific” (evaluation criteria are adapted to suit SSH specifics); “metric” (relying on metrics and data during the evaluation procedures); “performance-based funding” (better performance increases chances to get funded); “push for English” (requirements to increase the number of publications in English); “gender” (gender issues are taken into account, such as maternity leave); “career promotion” (institution that centrally organizes career promotion); “SSH grants” (specific calls for SSH research). Combining these categories with the countries’ data, five ideal types of national evaluation systems (NES) crystallized:

NES1: No national database, non SSH-specific: Cyprus, France, Iceland, Macedonia, Malta, Montenegro, Portugal, and Spain; with proximity of Bulgaria and Italy.

NES2: Non-metric, SSH specific: Austria, Germany, Ireland, the Netherlands, Serbia, and Switzerland.

NES3: Performance-based funding, non-metric: Lithuania, Norway, and South Africa; with proximity of Denmark and Israel.

NES4: Performance-based funding, metric: Czech Republic, Croatia, and Poland; with Finland in proximity.


This classification illustrates that the E – C – O – R scheme is obviously oversimplified. There is no single unit of C (assessment criteria), it has at least five rather different variations, some of which are more explicit regarding the criteria, others less. Furthermore, it is not only about criteria but also about methods: peer review or metric-based assessments are quite different even though metric assessments include aspects of peer review and vice-versa (see Ochsner, Kancewicz-Hoffman, Holowiecki & Holm 2020). Thus, instead of C we have to deal with C-NES1, C-NES2, C-NES3, C-NES4 and C-NES5. Having in mind the fact that six countries are only in proximity to the ideal types, there might be additional C-NES types. This classification gives a first hint that the E – C – O – R scheme has decreasing chances to withstand the reality check.

LINGUISTIC DIVERSITY IN SSH

One of the authors of this article is Lithuanian. And Lithuanians are famous for being obsessed by the national language. Thus this issue, on the one hand, cannot be avoided and, on the other hand, will receive disproportional attention. It is one of the prices of diversity.
Looking from the privileged perspective of native English speakers, the language of research is not an issue at all. Do your research, submit the results to any prestigious academic journal of your choice and wait for the peer review, and if the article is of high quality, you’ll get published. The outreach is guaranteed by the quality and status of the journal and not by the language. Every respectable scholar is capable of understanding the global international language and operating it sufficiently, isn’t she? Criteria in different disciplines might vary but the pattern itself stays intact.

Switching perspective from native English speakers to those SSH researchers, whose English is a second (or even third) language, it does not look that simple. The linguistic range of ENRESSH is approximately 30–32 languages plus English as the *lingua franca*. This means that, on a daily basis, the majority of network participants speak, teach and conduct research in languages other than English. In a truly globalized world, non-English speakers would definitely behave according to the same publishing pattern as native English speakers without any thoughts of publishing in a local language or dialect. But so far this is not the case. The actual publication reality reveals one of two things: either the world is not fully globalized, thus an enormous amount of research currently being presented in “local” languages is a temporary setback and sweet English is waiting for us; or the pattern is limited and cannot explain the behavior of “indigenous” researchers preferring vernacular languages, therefore the pattern does not work as a universal explanatory principle.

It does not matter which one of the two is correct, insofar as scholars in the SSH experience stronger or weaker, outer or inner, institutional or individual, but still constantly present pressure to publish in English. What do the researchers collaborating under the ENRESSH label have to say on this? The answer is obvious, i.e., *everything is diverse, thus*… It is time to take a closer look at some of the interesting data.

Publication data based on peer-reviewed articles presented in Kulczycki et al. (2020) reveals huge discrepancies in the research language preferences across countries. Based on the data from seven European countries, the study demonstrates two extremes: on the one hand, there is Norway, where 84.5 per cent of the articles are published in English, 13.1 per cent in Norwegian, and 2.3 per cent in other languages. On the other hand, data from Poland shows only 21 per cent articles in English, 74.2 in Polish, and 4.8 in other languages. Five other countries under consideration – Finland (respectively 78.3, 18.1, 3.6), Flanders (75.2, 20.6, 4.3), Denmark (69.7, 27.7, 2.6), Slovenia (46.8, 46.3, 6.9) and the Czech Republic (33.5, 58.8, 7.8) – are placed in-between Norway and Poland, where the first three linguistically are closer to the Norwegian publication pattern, the Czech Republic is closer to Poland, while Slovenia stands the middle ground with share of English.
and Slovenian being nearly equal. Looking at the data, one might be tempted to combine this data with the classification of high-performing and low-performing countries, which would result in a clear-cut recommendation for Poland and the Czech Republic to increase the number of publications in English, because English means quality and English is the language of high performance. Luckily, percentages alone are not sufficient to base value judgments on what academic publishing ought to look like linguistically.

The data above on linguistic diversity can be combined with the *Helsinki Initiative on Multilingualism in Scholarly Communication*[^3]. ENRESSH members actively participated in declaring and signing the initiative, which addresses different types of stakeholders (policy-makers, leaders, universities, research institutions, research funders, libraries, and researchers) in order to convince them to support and promote multilingualism in research during the processes of dissemination, publishing, assessing, evaluating, and funding:

… Research is international. That’s the way we like it! Multilingualism keeps locally relevant research alive. Protect it! Disseminating research results in your own language creates impact. Endorse it! It is vital to interact with society and share knowledge beyond academia. Promote it! Infrastructure of scholarly communication in national languages is fragile. Don’t lose it! … (*Helsinki Initiative on Multilingualism in Scholarly Communication* 2019).

**THE STAKEHOLDERS’ DIVERSITY**

Another important illustration of the ENRESSH-discovered diversity is the diverse range of stakeholders, who constantly participate in the SSH evaluation processes. In the initial common-sense explanation presented before (E – C – O – R), there is no place for a SH (stakeholder) category, which is obviously wrong. It is impossible to detach a SH category from the picture, because stakeholders are omnipresent and are found everywhere during the evaluation. The authors of *ENRESSH Policy Brief on Research Evaluation* (Ochsner, Kancewicz-Hoffman, Ma et al. 2020) even claim that, out there, one can find four major (SH1, SH2, SH3, SH4) and three intermediary (SH1-SH2, SH2-SH3, SH3-SH1) categories of stakeholders, which result in the following taxonomy of at least twenty different types of stakeholders:

**SH1. Research production:**
- SH1.1. Researchers
- SH1.2. Colleagues
- SH1.3. Learned societies

SH1-SH2. Intermediaries:
SH1- SH2.1. Researchers from other fields
SH1- SH2.2. Cultural institutions

SH2. Research consumption and use:
SH2.1. Business
SH2.2. Interested lay public
SH2.3. Media and cultural production
SH2.4. Taxpayers

SH2-SH3. Intermediaries:
SH2-SH3.1. Ministries
SH2-SH3.2. NGOs

SH3. Research policy and administration:
SH3.1. Research councils
SH3.2. University administration
SH3.3. Funders
SH3.4. Ministry of research and education

SH3-SH1. Intermediary:
SH3-SH1.1. Academies

SH4. Evaluation Services:
SH4.1. Bibliometricians
SH4.2. Data providers
SH4.3. Evaluators
SH4.4. Experts/Peers

Looks complicated? Definitely. Each stakeholder has its own interests. Criteria and preferences (weightings) between criteria might differ significantly among stakeholders. Therefore, in an evaluation, different interests have to be balanced out. The further we go referring to the ENRESSH finding, the more we perceive reality’s mercilessness to the E – C – O – R scheme. The taxonomy of multiple stakeholders demonstrates the scheme being an oversimplified and indeed a naïve one. All the above reveal that the E is not independent, i.e., she depends not only on one of the C-NES, but on a certain number of stakeholders as well. And that presupposes different national contexts together with diverse linguistic realities.

DIVERSITY IN IMPACT CREATION

One of the novelties for ENRESSH members from a widening number of countries was the impact discourse, which signals that even in SSH the societal impact is imminent. One can guess that the impact might be intentional, as well as accidental.
or unintentional. But, according to Muhonen et al. (2020), the social impact creation (IC) in SSH implies at least four general (or strategic) models, which are divided into twelve detailed (or tactical) models (Muhonen et al. 2020: 42):

IC1. Dissemination:
  IC1.1. the interactive dissemination model

IC2. Co-creation:
  IC2.1. the collaboration model
  IC2.2. the public engagement model
  IC2.3. the expertise model
  IC2.4. the mobility model

IC3. Reacting to societal change:
  IC3.1. the ‘anticipating anniversaries’ model
  IC3.2. the ‘seize the day’ model
  IC3.3. the social innovation model
  IC3.4. the commercialization model

IC4. Driving societal change:
  IC4.1. the research engagement as a key to impact
  IC4.2. the knowledge ‘creeps’ into society model
  IC4.3. the building ‘new epistemic communities’ model (Muhonen et al. 2020: 42).

Along with the addition to the initial evaluation schemes, there is a need to take into consideration an IC factor, i.e., to determine in which intentional or unintentional ways the R via O participates in IC.

OTHER UNAVOIDABLE DIVERSITIES

Publication diversity in SSH. ENRESSH researchers reveal different publication patterns as well as different linguistic publication patterns. An ENRESSH-based comparative study of publication patterns in eight European countries (Kulczycki et al. 2018) shows that the SSH field is diverse in its publication type preferences. For example, the percentage of articles among other types of publications in Slovakia (77.1%) significantly exceeds the percentage of articles in Poland (44.8%); or the percentage of monographs in the Czech Republic (12.3%) is nearly 7 times higher than in Flanders (1.8%), or approximately 4 times higher than in Finland (3.3%) and Norway (3.4%), etc. (ibid). The deceitfully small percentage of books in overall scholarly production does not mean a decline of scholarly books; quite the contrary, “book publications are and remain vital for the SSH” (Engels et al. 2018: 603).
Gender diversity in SHH. Searching for further diversities, Guns et al. (2019) explored the territory of gender differences in journal publishing in SSH. Most SSH disciplines are dominated by men, except for Education, Psychology and Cognitive Sciences, and Languages and Literature, which have a majority of female researchers. Disciplinary gender differences do not translate into differences in collaboration rates, as women tend to collaborate at similar rates as men. Female authors publish fewer articles but they are more likely to publish in a local language than their male colleagues. However, these differences are small (Guns et al. 2019: 69–71). However, women seem to react differently to evaluation than men; women try to adapt their work to the (anticipated) evaluators’ judgment even when they disagree and think that their own ideas might not be less interesting or correct, while men form their own identity by defending their ideas and approaches, taking up good advice but refusing points about which they disagree. This might potentially lead to a disadvantage in career development for women (Lendák-Kabók & Ochsner 2020).

Summing up, along with multiple C-NES, SH, IC, disciplines, and languages, these additional diversities also cannot be ignored and, therefore, adding them to the initial evaluation scheme makes any abstract but comprehensive evaluation model nearly impossible.

WHAT’S NEXT?

But what do all these uncovered diversities teach us? What do we do with this information? What are the next steps? During the ENRESSH-related RESSH conference in Valencia 20194 Jordi Molas-Gallart asked: OK, everything is diverse, there is diversity everywhere, diverse this, diverse that, but what to do with the fact that everything is diverse?

If we think about this diversity creatively, the fact of diversity contains some hidden potential. Which is ... thinking outside the box and combatting oversimplified but still influential interpretations of the SSH research evaluation. The first step was taken and resulted in the possibility to gather people representing different countries, disciplines, ages, genders, experiences, languages, and social roles related to various engagements of research evaluation: the evaluated and the evaluators, researchers and stakeholders, administrators and policymakers.

What can be more mundane in a researcher’s life than evaluation? At first sight, it might seem there is nothing more routine than scholarly evaluation. Every active

researcher constantly participates in the process of evaluation either as an evaluator, as the one being evaluated, or both. Usually the evaluation is perceived as a natural and inseparable part of scholarly life. She peer-reviews the texts (books, articles, project proposals, reports, etc.) of her colleagues, she approves metrical data, etc. Simultaneously, she is at the other side of the evaluation procedures as her own scholarly output is evaluated. She is both a subject and an object in a continuous research evaluation process. As simple as that. Almost everyone does that, but only a few ask what evaluation is. What does it mean to evaluate? What is the meaning of “evaluation”? What is the aim of evaluation? How do we know that any particular research is of any value? How are values created in the evaluation process? Do evaluators recognize intrinsic values of research, or do they rather attribute extrinsic values to it? What is quality in SSH research?

In fact, what is the use of a simple and standardized research evaluation procedure? Research is itself multifaceted and diverse. Researchers and academics in general should know: there is no single truth or, at least, it is not fully accessible to the human mind. All we can generate is knowledge that is context-specific. Of course, the contexts vary in scope; some research might be generally valid while other research might be valid in a very specific context. Therefore, evaluation needs to take the context into account. Even more so if we accept that evaluation is not about truth or not-truth but rather about assigning value. Then, it becomes obvious that value is not constant over time and place. Diversity is thus not a problem that makes evaluation difficult and almost useless (as it is not comparable across different contexts) but it is the very essence of evaluation. Only when put in the appropriate context, can research be adequately evaluated. If we want to take out most of research, its evaluation needs to be led by the principle of diversity.

Therefore, ENRESSH’s mission in D-Land is by no means finished. On the contrary, it might be never-ending; thus, it is dangerous to mess with diversity, for otherwise it would be nearly impossible to understand evaluation properly. And, to paraphrase a famous Voltaire quote, if there were no ENRESSH, it would be necessary to invent it. The best is yet to come.

ACKNOWLEDGEMENTS

This article 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).
REFERENCES


Aldis Gedutis, Michael Ochsner

ENRESSH: MISIJA Į KRAŠTĄ D


RAKTĄŽODZIÁI: ENRESSH, socialiniai mokslai, humanitariniai mokslai, mokslinių tyrimų vertinimas, skirtingų, įvairių, nacionalinės kalbos, socialinis poveikis.