

Towards research assessment reform for the social sciences and humanities: progress and insights from the GraspOS project¹

Fotis Mystakopoulos

ORCID: <https://orcid.org/0000-0002-9354-3754>

OPERAS Research Infrastructure, Belgium

fotis.mystakopoulos@operas-eu.org

Carol Delmazo

ORCID: <https://orcid.org/0000-0002-8284-9117>

OPERAS Research Infrastructure, Belgium

carol.delmazo@operas-eu.org

Libellarium 16 (1) 2025: 3–26

Original Scientific Report / Izvorni znanstveni rad

UDK: 001.818

Received / Primljeno: 22. 09. 2024.

Accepted / Prihvaćeno: 24. 10. 2025.

doi: <https://doi.org/10.15291/libellarium.4548>



Abstract

Purpose. This study focuses on developing research assessment criteria for the Social Sciences and Humanities (SSH) through the GraspOS project. The aim is to reform traditional evaluation methods that rely heavily on bibliometric indicators and to produce recommendations tailored to the SSH community, addressing the limitations of current practices that predominantly emphasise quantitative metrics.

Approach/methodology/design. This study presents outcomes from three workshops with participants from various SSH backgrounds. The workshops facilitated discussions on values-driven assessment, output diversity, and the integration of qualitative indicators. The data collected were analysed using the deductive method, with a priori codes established.

Findings. The analysis revealed consensus on the need for a more inclusive evaluation framework that recognises diverse outputs beyond traditional journal articles. Participants emphasised the importance of qualitative measures, such as peer review, and proposed using narrative CVs to better reflect individual contributions.

Research limitations. The study was conducted within a specific context of the Responsible Research Assessment (RRA) movement and the GraspOS project, which may not fully

represent the broader SSH landscape. Generalisation to all SSH disciplines is not possible. **Practical implications.** The study provides recommendations for formulating generic research assessment criteria as suggested by SSH scholars. **Originality value.** This work provides valuable insights into developing more equitable and effective assessment practices in academia for the SSH community.

KEYWORDS: open science, research culture, research evaluation, responsible research assessment, social sciences and humanities

1. Introduction

The reform of research assessment system has been ongoing for over a decade now (Curry et al. 2020; de Rijcke et al. 2023). This primarily involves reducing reliance on bibliometric indicators as proxies for quality and impact during evaluations². In their 2020 study, Curry et al. identified fifteen key initiatives to facilitate this transformation, which they labelled “movers and shifters”, all associated with the concept of Responsible Research Assessment (RRA), which is another term used to describe the ongoing reform (Curry et al. 2020).

The European Commission (EC) is investing in alternative approaches for research assessment that embrace Open Science. As evidenced in a recent EC report, GraspOS was one of thirty Horizon 2020 and Horizon Europe projects addressing research assessment reform as outlined in the Agreement on Reforming Research Assessment (European Commission, European Research Executive Agency, Oancea, and Wilson 2024). The number of projects and the scale of funding demonstrate the complexity and magnitude of the task. The GraspOS project is referenced in the report on research assessment across different clusters developing frameworks, as it developed the SCOPE+i Framework, originally labelled the Open Science Assessment Framework (OSAF)³ (Tatum et al. 2023).

SCOPE+i is part of the federated infrastructure that GraspOS developed. The infrastructure offers opportunities for monitoring and exploring data sources, while also providing a framework with guidelines on responsible use of data and tools in research assessment. The GraspOS Project employed nine pilot studies to test the various functionalities and provide feedback for the developments of templates and recommendations for SCOPE+i.

The pilot studies with researchers from Social Science and Humanities (SSH Pilot) were led by the OPERAS Research Infrastructure⁴. One of the key objectives of the pilot was to pro-

² In this manuscript, the terms *evaluation* and *assessment* are used interchangeably. This reflects their common usage in the scholarly and policy discussions that informed our work, and no conceptual distinction between them is intended.

³ Although the project initially developed its resources under the name Open Science Assessment Framework (OSAF), subsequent internal discussions led to a renaming. The revised title, SCOPE+i, was adopted to acknowledge both the influence of the SCOPE Framework for Research Evaluation, which underpins the project’s methodological approach, and the additional dimension denoted by +i, referring to infrastructure. To access SCOPE+i webpage: <https://graspos.eu/open-science-assessment-framework>.

⁴ OPERAS Research Infrastructure. <https://operas-eu.org/>

duce generic research assessment criteria tailored to the needs of the SSH community⁵. As a first direct action to engage with this community, a GraspOS Community of Practice (CoP) was held in January 2024 to discuss the role of Open Science in fostering a more inclusive research assessment in SSH (Mystakopoulos 2024). During the CoP there was discussion about the diversity of research outputs, and what could or should be included in a research assessment exercise. For example, there was a reference to the development of the concept of non-traditional outputs in Australia (Australian Research Council 2019). Another topic raised was the limitations of the Contribution Role Taxonomy (CRedit), among others. This event shaped the initial ideas for how to run the workshops that took place within the SSH Pilot.

2. Literature review

2.1 SSH and Research Assessment

Research assessment in the field of SSH is a contentious issue that can generate considerable debate (Ochsner, Hug, and Galleron 2017; Reale et al. 2017). There are methodological perspectives on how to assess the quality, impact and excellence of research (Galleron et al. 2017); questions arise about the databases used to assess research, and how SSH is affected by these choices (Archambault et al. 2006; Sivertsen 2016); and both issues lead to discussions about the benefits and limitations of using scientometrics or bibliometrics in research assessment, a topic often broader than SSH (Leeuwen 2013; de Rijcke et al. 2023).

Reale et al. (2017) highlight that the need to monitor public spending in relation to investment for scientific research has led to the development of bibliometrics, scientometrics, and infometrics. There is consensus that research assessment from a bibliometric perspective, a quantitative approach, originates in the STEM (Science, Technology, Engineering and Mathematics) fields, and has since extended to the SSH (Archambault et al. 2006; Galleron et al. 2017; Leeuwen 2013; Ochsner, Hug, and Galleron 2017; Reale et al. 2017). Assessment exercises frequently use the Journal Impact Factor (JIF) or the h-Index (Reale et al. 2017) as proxies for impact and quality. Aksnes et al. (2019) discuss how quantitative indicators, particularly citation-based measures, offer advantages such as scalability, objectivity and the ability to efficiently capture certain aspects of scientific impact; however, these indicators also have well-documented limitations, as citations reflect only some dimensions of research quality and are influenced by disciplinary norms and database coverage (Aksnes et al. 2019). Both JIF and h-Index disadvantage SSH researchers for several reasons: SSH has a more diverse scholarly output than the STEM (Archambault et al. 2006; Hicks 2005) and this diversity is not adequately represented in the main assessment databases, which are SCOPUS and Web of Science (Leeuwen 2013; Ochsner, Hug, and Galleron 2017; Sivertsen 2016). To understand why SSH outputs are underrepresented, it is important to consider the types of diverse outputs identified in the SSH literature. Archambault et al. (2006) de-

⁵ See GraspOS: next Generation Research Assessment to Promote Open Science: <https://graspos.eu/social-sciences-humanities-domain>

scribe SSH's diverse portfolio of scholarly publications which extends journal articles. For example, SSH, has a higher proportion of book publications; citation patterns also differ, as SSH outputs typically receive fewer citations in the short term but accumulate value over longer periods due to extended citation lifecycles, with books often remaining relevant and cited over extended periods (Sneijder 2019). SSH research also has a local orientation; it can be locally relevant, and publications can be published in national languages or national journals not indexed in the major databases such as Scopus and Web of Science (Ochsner, Hug, and Galleron 2017). On the same topic, Hicks (2005) summarises four types of literature of SSH: journal articles, books, national literatures, and non-scholarly literature. These are some of the fundamental issues related to the bibliographic databases currently used for assessment. The problem became more pronounced when Scopus and Web of Science became the standard tools for assessing researchers and research, focusing on journal articles and their associated infrastructure (Sivertsen 2016). Overall, the ongoing challenge is that the pursuit and reliance on standardised bibliometric tools such as the Social Sciences Citation Index (SSCI) for research assessment may undermine the diversity and richness of the scientific literature in SSH (Hicks 2005).

Another point to consider in the context of SSH research assessment and the broader discussion of RRA is the need to strike a balance between qualitative and quantitative criteria (Reale et al. 2017; Schmoch et al. 2010; Wang 2023). Reale et al. (2017) briefly discuss the concept of assessing processes, emphasising the importance of evaluating how results are produced rather than focusing solely on outcomes. Schmoch et al. (2010) examine the various activities of researchers, highlighting the significance of contributions to scientific infrastructure, graduate teaching, and publishing. They suggest a more comprehensive set of indicators to assess scientific performance.

We conclude this section with two sets of recommendations regarding the future of research evaluation in SSH (Galleron et al. 2017; Ochsner, Hug, and Galleron 2017). Firstly, Galleron et al. (2017) advocate for a deeper understanding of the diverse outputs of SSH research and the involvement of a broad range of stakeholders in the evaluation process, including those being evaluated. It is important to consider how knowledge is applied in real-world contexts when evaluating research, as this is central to SSH. A more comprehensive evaluation is needed, balancing academic excellence with societal relevance and diverse responsibilities (Galleron et al. 2017). In addition to the proposals from Galleron et al. (2017), another set of recommendations aims to refine research evaluation. According to Ochsner, Hug and Galleron (2017) future evaluations should consider informed peer review as a qualitative measure, involving all parties in the review process. SSH researchers should play a central role in developing a wide range of qualitative criteria, which, where possible, should be defined by measurable indicators. Based on these agreed criteria, an evaluation sheet should be developed that includes both measurable and non-measurable elements. Non-academic criteria can also contribute to the evaluation process, provided their source and purpose are transparent. Each criterion should be assessed independently to ensure a more accurate reflection of performance. Finally, overall rankings or ratings should not be published. Instead, the focus should remain on the individual criteria, ensuring a more nuanced assessment (Ochsner, Hug, and Galleron 2017).

2.2 Responsible Research Assessment (RRA) and GraspOS

The RRA movement was the operational framework of GraspOS. Therefore, we present a definition of RRA as articulated by Curry et al. (2020, 4):

“This is an umbrella term for approaches to assessment which incentivise, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures”.

Several elements that appeared in the SSH section affect the broader scientific community. The wider RRA movement is trying to address misguided incentives regarding what kind of research can and should be funded, or what research is considered valuable, based on very specific criteria and indicators (Curry et al. 2020; Himanen et al. 2024; de Rijcke et al. 2023). The vision is to make research culture more open and inclusive in the future (Curry et al. 2020). Among the various initiatives related to RRA, a few are particularly relevant to the context of GraspOS: the Humane Metrics Initiative (Agate et al. 2022); the SCOPE Evaluation Framework (Himanen et al. 2024); the Coalition for Advancing Research Assessment (CoARA)⁶, and the Barcelona Declaration⁷. In relation to the idea of *processes* (Reale et al. 2017), the Humane Metrics Initiative is ideologically aligned with the values of SSH (Agate et al. 2022). The focus of this initiative is to identify new ways to assess SSH scholarship from a values-driven perspective (Agate et al. 2022). Key values associated with this initiative are equity, openness, collegiality, soundness and community⁸. HumetricsHSS is also cited as an inspiration for SCOPE (Himanen et al. 2024). SCOPE is a five-stage model for responsible evaluation (Himanen et al. 2024). The acronym stands for Start with what you value, Context Considerations, Options for evaluating, Probe deeply, and Evaluate your Evaluation⁹. SCOPE is also based on three key principles: evaluate only when necessary; evaluate with those being evaluated; and draw on evaluation expertise (Himanen et al. 2024). Both HumetricsHSS and SCOPE can be implemented through workshops, hence their relevance to the SSH pilot.

GraspOS was aligned with and supported CoARA and the Barcelona Declaration. In relation to CoARA, signatories to the Agreement on Reforming Research Assessment (ARRA) (CoARA 2022) commit to prioritising peer review as part of their assessment practices, supported by the responsible use of indicators. Like the San Francisco Declaration on Research Assessment (DORA, 2012), CoARA encourages organisations to move away from the bibliometric indicators that have dominated evaluations to date and to avoid using of rankings in assessment. This aligns with the approach exemplified by Utrecht University and the University of Lorraine, as previously outlined in the introduction. Finally, CoARA proposes that organisations consider the diversity of outputs and careers within research organisations when conducting research assessment. Critics of CoARA note a lack of understanding regarding the appropriate application of scientometrics (Abramo 2024; Ioannidis and Maniadis 2023).

⁶ See more details on the Coalition for Advancing Research Assessment here: <https://coara.eu/>

⁷ The full name of the declaration is: Barcelona Declaration on Open Research Information – Find the commitments here: <https://barcelona-declaration.org/>

⁸ See details on the Humane Metrics Initiative here: <https://humetricshss.org/our-work/values/>

⁹ Find more details about the SCOPE here - <https://inorms.net/scope-framework-for-research-evaluation/>

They also point out that peer review may not be a sustainable approach in the long term (Abramo 2024). Essentially, they argue that scientometrics offer excellent value for money when applied correctly and provide opportunities for assessing scientific outputs in ways that peer review cannot (Abramo 2024; Ioannidis and Maniatis 2023). Similarly, Aksnes et al. (2019) emphasise that, while peer review provides contextual judgement and qualitative insight, it still faces challenges related to workload, subjectivity, sustainability, and scalability. In response, supporters of the CoARA values point out that the commitments do not exclude the use of quantitative methods but instead emphasize the involvement of expert assessment (Balboa et al. 2024). This debate serves as a reminder of the difficulty of balancing between quantitative and qualitative assessment measures. It is also important to mention the Barcelona Declaration, initiated relatively recently. The Declaration calls for signatories to prioritise the use of open research information.

One example is the recent commitment by the University of Sorbonne to support OpenAlex, an open alternative to the commercial and proprietary Web of Science database. Most recently, open science monitoring was highlighted as an area of concern by the UNESCO report on Open Science outlook (Rafols 2024; UNESCO 2023). The argument is that by relying on easily monitored sources, such as journal articles, there is a risk of excluding a significant portion of research activity that cannot currently be monitored. To include diverse processes and outputs in research assessment, they should, to some extent, be measurable (Ochsner, Hug, and Galleron 2017). GraspOS delivered its contribution to this issue in the form of data, tools and services (Vergoulis et al., 2024). Widening the scope of evaluation aligns with the goals of SSH, as identified in the previous section, which highlighted the necessity for diversity in the production of knowledge of research in SSH and that any evaluation should consider not only outputs but also processes and other types of activities.

2.3 Research questions

This literature review was conducted to provide an account of the main topics found in the discussion of research assessment and SSH. It is not intended to be exhaustive, but rather to provide a snapshot of the current debate. In terms of originality, this study presents a collection of thoughts and statements from participants in the consultation workshops, which explored these debates through specific exercises. As representatives of the SSH pilot, our objective is to facilitate a discussion on the optimal methods for representing the impact and performance of SSH research in evaluation panels. We aim to propose alternative approaches that go beyond quantitative measurement and align more closely with the unique practices and voices of the SSH community, as expressed during the project. As explained in the introduction, we conducted consultation workshops with the SSH community. Based on the information available from the literature, we arrive at the following research questions:

1. What values should guide the assessment of research in the SSH?
2. What outputs, processes, and practices should be considered in the assessment of research in the SSH?
3. Which indicators are important or missing in the assessment of research in the SSH?
4. What platforms, tools, and services should be used to assess research in the SSH?

5. From an open science perspective, what should be prioritised in the assessment of SSH research?

From the research questions, we also established several research objectives:

- To explore the challenge of balancing quantitative and qualitative criteria in SSH research assessment.
- To document the diversity of research outputs and activities in SSH and how they can be incorporated into assessment frameworks
- To determine the core values and priorities that should shape responsible research assessment practices in SSH.
- To assess the role of digital platforms, tools, and services in supporting the research assessment process for SSH scholars.

3. Methodology

The methodological origins of this research can be found in the SCOPE Research Evaluation Framework (Himanen et al. 2024) and in the materials produced by the HumetricsHSS initiative workshop material¹⁰. To gain insights into how current researchers experience research evaluation, we conducted workshops in line with the recommendations outlined by SCOPE and the HumetricsHSS initiative. The process was designed to facilitate an exploration of participants' views about values, processes and current practices in research assessment, with the ultimate objective of documenting recommendations for improvement. From SCOPE's perspective, one of the three key principles is to evaluate with those being evaluated. The Humane Metrics Initiative is more closely aligned with the SSH domain and recommends going beyond the discussion of the research outputs but also to consider processes and practices in evaluation. Three consultation workshops (CW) were held between 13 March and 9 April 2024, with a total of 18 participants. The following sections present details of these consultation workshops. Detailed methodological documentation, including the structure and organisation of all workshop exercises, is provided in an accompanying Zenodo deposit,¹¹ to ensure transparency and to facilitate possible reuse of this approach.

3.1 Sample population

To maximise diversity in research roles and countries of employment, we organised three workshops, each with a minimum of five and a maximum of eight participants. This ap-

¹⁰ The Humane Metrics Initiative produce a workshop kit that can be reused. However, it still incomplete, so please use with caution - <https://humetricshss.org/your-work/workshop-kit/>

¹¹ Full citation: Mystakopoulos, F., & Delmazo, C. (2025). Supplementary Methodology Documentation: Consultation Workshops for Research Assessment Reform in the Social Sciences and Humanities (GraspOS SSH Pilot). Zenodo. <https://doi.org/10.5281/zenodo.17755599>

proach ensured that all participants had sufficient time to express their thoughts and opinions during the discussions. To recruit participants, we issued a public call through various communication channels: OPERAS and GraspOS social media accounts (LinkedIn, X), multiple mailing lists, and a public blog post on the OPERAS Hypotheses platform. We successfully recruited participants from a range of countries representing diverse cultural and research backgrounds, thereby creating a culturally responsive focus group (Rodriguez et al. 2011). Although we do not analyse demographic variables in detail, we ensured representation from a broad spectrum of European countries and professional contexts. Participants were based in Northern, Southern, Western, Central, and Eastern Europe, with no region receiving disproportionate attention or participation. The group also included individuals at various academic career stages (from doctoral researchers to senior scholars), those with expertise in research assessment, open science experts, and library and research support professionals.

3.2 Ethical considerations

Throughout the process, we adhered to the guidelines on informed consent as outlined by Krüegel (2019), ensuring that participants were aware of how their data would be handled and to guarantee that their privacy would be safeguarded. To ensure the confidentiality of the participants, and adhere to ethical guidelines, we implemented a structured anonymisation protocol that closely follows the guidelines established in the relevant literature and best practices for anonymising qualitative data (Information Commissioner’s Office 2021; Stam and Diaz 2023). Our approach drew heavily on the procedures outlined in FORS Guide No. 20¹², which provides theoretical and practical considerations for anonymising interview transcripts. The main issues relate to direct and indirect identifiers, specifically personal names and surnames as well as places of occupation and other background and professional associations (Stam and Diaz 2023). For names, we used aliases in order to maintain a semblance of personas, following the recommendation that aliases reflect similar background (Stam and Diaz 2023). For statements about the place of work, including locations, we categorised this information by aggregating it to a more abstract value, for example instead of using University of Southampton, the replacement would be *a University in the UK*.

3.3 Data Collection

To gather data, we opted for online Zoom recorded consultation workshops. Each consultation workshop was initially scheduled for 90 minutes, with a few additional minutes allowed in each workshop for closing statements and wrapping up of discussions. In each case we asked for the participants consent before proceeding to extend the time. The workshops were conducted as follows: a first round of short presentations of each participant and two ice-breaker questions; a presentation summarising the key objective of the workshop and how it was inserted in the context of the GraspOS project; and two exercises, each focusing on a specific topic, with approximately 40 minutes allocated to each topic.

¹² Access the FORS (Swiss Centre for Expertise in The Social Sciences) Guide here: https://forscenter.ch/wp-content/uploads/2023/03/qualitative-data-anonymisation_final.pdf

The three workshops were intentionally designed to, collectively, address all five Research Questions (RQs) guiding the study. Workshop 1 (CW1) covered RQ1 and RQ2 (values and SSH-specific outputs and practices), Workshop 2 (CW2) covered RQ3 and RQ4 (indicators and digital platforms), and Workshop 3 (CW3) addressed RQ5 by focusing on similar concepts to Workshops 1 and 2 but from an open science perspective.

Table 1. Summary of topics that were explored during each consultation workshop. Every workshop explored 2 topics, for approximately 40 minutes each

Consultation Workshop	Topic 1	Topic 2
Consultation Workshop 1 (CW1)	Values to guide research assessment in the SSH	Outputs, Processes and Practices to be considered in SSH research assessment
Consultation Workshop 2 (CW2)	Relevant indicators to assess research in SSH	Platform, tools and services suitable for SSH research assessment
Consultation Workshop 3 (CW3)	Open Science outputs, processes and practices to be valued in SSH research assessment	Indicators for measuring Open Science outputs, processes and practices in SSH

Some thematic areas appear in more than one workshop, and this overlap is intentional. Topics such as outputs, practices and indicators were first explored from a general SSH perspective (Workshops 1 and 2) and later revisited through an Open Science lens (Workshop 3). The purpose was not to duplicate content but to examine similar themes from different angles, ensuring that the workshops collectively covered all five research questions, while allowing participants to elaborate on SSH-specific and Open Science-specific considerations.

The transcripts of each recording represent the primary research material. We also provided digital boards in the form of Google Jamboard and Miro for the exercises, thereby facilitating the gathering of further information. These served as supplementary material and were consulted when necessary to facilitate the direction of the conversation.

3.4 Data Analysis

For the analysis, we organised the data using the deductive method, applying a priori coding categories informed by the research questions and objectives of this study (Bingham 2023), as well as prior literature and theory on research assessment and SSH. We did not use any theoretical framework for coding. Additionally, we examined the transcripts and coded responses by theme, and the analysis presents a narrative for each theme.

3.5 Study limitations

It is important to acknowledge that this study is situated within a highly constrained and limited perspective of the RRA movement, particularly in the context of the GraspOS project. We don't intend to present the findings as representative of the SSH domain, nor do we offer a generalisation of the results.

4. Analysis

The workshops aimed to explore particular concepts, all from the SSH perspective: (1) values-driven research assessment (2) diversity of outputs, processes and practices (3) indicators for research assessment (4) digital platforms for monitoring and reporting of activities and (5) the open science dimension of outputs and indicators.

4.1 Values-driven research assessment

During the internal exercise for GraspOS, on behalf of the SSH pilot we hypothesised what the key values for research evaluation in the SSH would be. The framework follows the SCOPE recommendation of organising the values into super-values, values and sub-values:

Super-values:

- Transparency
- Output Diversity
- Quality

Value:

- Inclusivity of diverse outputs and practices

Sub-values:

- Widen the scope of research outputs considered important for SSH Research Assessment: monographs; non-peer review journals; multilingual content; peer review practices (identify which sources can contribute to this diversity)
- Raise the profile of publication practices such as editorial or peer review participation (produce indicators that recognize and reward peer review)
- Create assessment protocols that respect both the diversity of outputs and diversity of practices.
- Provide support in the use of services that enable SSH monitoring within OPERAS RI and beyond: i.e. PRISM, Operas Metrics, GoTriple, Books Analytics Service

These values were presented to the participants and the subsequent discussion focused on the key ‘super-values’ proposed to guide research evaluation.

Transparency was well received and there was little objection from participants in each inclusion as a super value. However, there were comments of caution on what transparency entails as a definition and how to be used in practice. One participant referred to the transparency definition comparing different contexts such as hiring for new position, or transparency in the form of publishing:

“Depending on the kind of what associations exist around a super value such as transparency in different contexts. Where one might associate transparency as recruit or promotion processes where the board of an institute, the prospective person to actually get a new po-

sition and the committee for the evaluation share the reports with another. That's one way of ensuring transparency in the civility process. Another is, but sometimes just connected to ideas around dissemination of information, Open Access." - **Costas**

The topic of transparency led to discussions on the question of whether it is on equal footing with openness. There is a connection between transparency and openness in the sense that sharing data openly can result in greater transparency. However, transparency is not always about openness in the sense of accessibility, but rather there are layers to it, as noted in the discussion.

"Just to be precise in terms of practices, some practices might be transparent to our field or to the researchers, because we share, but they are not open because they can't be open, this is the case in sociology often. We aim to be transparent, but it doesn't mean we have to be open to be transparent." - **Zuzana**

Regarding questions on output diversity, there were concerns about maintaining the current focus on outputs when considering super-values in more detail. We therefore decided to remove the word *outputs* and maintain the term *diversity*, as this would accommodate a wider range of concepts including diversity of careers, languages, and outputs, which are crucial for SSH.

"But I wonder whether output diversity is really a super value or whether it's rather a reflection of the super values...Then maybe it's better to call it diversity because that would also include diversity in languages or diversity in career positions etc, as a super value that defines what we are after rather than having a very specific super-value."

- **Patrick**

Quality was not discussed in depth when exploring the super-values. However, it was raised during the discussion of the value statement and how it was initially absent. Several participants noted that the main issue with quality is how difficult it is to measure. Beyond this, the value statement was not further challenged, partly due to the nature of the exercise. The SCOPE framework should be applied in specific contexts, such as a university or a research department.

During the sub-values discussion participants expressed their thoughts on improving the four statements and suggested new ones. They added possible sub-values related to the topic of each original statement. For example, option 1 (see Table 2) urges us to look beyond what is measurable or tangible when diversifying outputs for inclusion. Option 2 suggests the need to slow down and publish less rather than maintaining the current pace of publication.

Table 2. Overview of the responses from the participants during Consultation Workshop 1, in relation to what they would like to see in the sub-values statement of a Values Framework for the SSH

Sub-value statement
1. Go beyond measurable, tangible. Community engagement through teaching. Transdisciplinary research practices.
2. Publish less, not more. Taking into account the individual role in research processes.
3. No bean counting, rather plans: what did you do to reach what?
4. Evaluation to the top: What did you miss (from superiors, institute etc.) to reach your goals?

Overall, while there was discussion about how each value could be interpreted in different contexts and how the initial super-values, values and sub-values statements could be improved, participants seemed to agree on the sentiment these values represent in relation to the objective of SSH.

4.2 Diversity of outputs, processes and practices

A fundamental discussion regarding the evolution of research assessment has focused on how to increase diversity in what is included in a research evaluation, specifically how to move beyond the publication of journal articles as the sole measure of success.

It was acknowledged that certain outputs continue to have a dominant influence on the research assessment process. For example, researchers must include their monographs and research articles to be considered for hiring and promotion. However, participants emphasised that the assessment process should also consider other outputs, such as the development of databases, software, questionnaires or surveys, games and interactive programmes or objects.

The activities undertaken by the researchers were also discussed. From a PhD perspective, one participant highlighted the necessity of attending conferences, but not every conference. He explained that it's also important to consider who organises the conference and whether it is a recognised organisation. Prestige appears to be important not only in journals, but also in conferences. However, other activities beyond conferences should also be considered. Work in scientific organisations such as committee or editorial duties could be taken into account, as well as peer review activity, providing assistance to colleagues with research methods, supervision of junior colleagues, interactions with policy, society and Non-Governmental Organisations (NGOs), and other forms of input, regardless of whether they result in an immediate outcome, are some examples mentioned.

It was also noted that given SSH aims to have an impact on societies, that impact ought to be measured. Societal impact takes the form of engagement with a diverse range of stakeholders resulting in activities that are not always obvious or easily fit in the academic structures.

“And also, I don’t know, for instance, in societal contribution, I think that especially when we talk about social sciences and humanities, it shouldn’t be more clear that it’s also a contribution to a cultural life or to culture. I don’t know exactly, but for instance, contributing to translations or to cultural discussions or to writing critiques of cultural art objects, art events or something like that. So that kind of contribution should also be valued.” - Petra

4.3 Indicators for research assessment in the SSH

While CW1 was a conversation about outputs and how they could be more diverse, CW2 aimed to discuss how to measure different outputs, how to turn these diverse outputs, and other activities or processes beyond scientific publication into indicators. Often, when discussing indicators for assessing research in the SSH, the conversation seeks to understand how a balance between qualitative and quantitative indicators can be achieved.

There was universal agreement that the H-index and its derivatives should not be used when assessing SSH output. In CW2, when discussing the topic of indicators as proxies for quality, it was again noted that quality is difficult to measure. Thus, doubts about how to measure quality arose in both workshops. One of the participants responded to the quality issue by suggesting the inclusion of peer review measures as a qualitative indicator. There was agreement that peer review activity is an indicator of quality:

“Thus, I propose to add peer reviewing as a qualitative indicator.” - Chiara

“So always this kind of assessment by peers is something that should be the basis of some kind of quality assessment, I think.” - Petra

However, participants also discussed the limitations of assessing the quality of the peer review input, and how peer review currently lacks standardisation, thus alluding to differing standards of what constitutes a good peer review. The advent of Generative Artificial Intelligence (AI) was also discussed, along with its pitfalls, evidenced by examples in the literature where the use of AI has led to the publication of questionable articles. Therefore, establishing peer review as a proxy for quality becomes a practice that needs to be examined as suggested by SCOPE for possible cases of misuse and manipulation. As a countermeasure, open peer review was suggested. While this line of discussion has proved insightful, it clearly demonstrated the need for further exploration of peer review as a qualitative indicator.

It is also important here to note that indicators, and the direction of measurement by quantitative indicators of SSH outputs, even if more diverse outputs and practices were explored, also met with resistance during CW1:

“But I like the comment to go beyond measurable, and I have a problem producing new indicators in general. Would they really make a difference? I don’t really think so. They would just make another system that will kind of exclude people that do not fit in the system that is measured. But this is not about values that much, this is more about how do we realise those values and put them into practice.” - Zuzana

This comment responded to the idea that diversifying outputs and creating more indicators does not address the fundamental issue of how assessment is being performed at the moment which is by “counting beans” method, which is trying to always measure in pure numbers what the contribution of each person is in different key performance indicators.

4.4 Digital platforms and monitoring and reporting of activities

A fundamental aspect of the project revolved around the tools and services in the form of digital platforms, effectively being the technical aspect of the infrastructure. The discussion during CW2 focused on one of the services offered by Athena Research Center: BIP! Scholar¹³(Vergoulis et al. 2022). BIP! Scholar is a platform that allows researchers to create profiles representing their research activities, enrich them with contextual information and indicators, and highlight different aspects of their research careers. The main objective of the platform is to offer profiles that cover a wide range of research activities, going beyond scientific publications (Vergoulis et al. 2022). Initially, features of the platform were presented as a solution to offer tailored indicators for SSH researchers. We encouraged participants to employ a “blue-sky”¹⁴ thinking and to suggest how this service could be of use to them. Initially, the reaction was that the interface and the focus of the service was leaning towards quantitative representation of a researcher’s profile. This is in contrast with the values and expectations of SSH scholars.

“But generally, my impression is that this is really too quantitative to be appropriate for SSH.” - Petra

“This kind of qualitative indicators are tailored to STEM, they are far from our habits, and this is immediately what you feel when you read this.” - Chiara

“[...] it is still very, very quantitative, and very, very far from our behaviour, disciplinary behaviours. So, I don’t think that this can be accepted in a humanities context.” - Chiara

Another comment regarding the technical aspect of the service is whether the data are interoperable and if it can communicate with established identifiers, such as ORCID profiles, and whether it could be linked with national Current Research Information Systems (CRIS) where a lot of research in the SSH may be documented.

“What I would like to see and understand better is how this tool is going to be interoperable and interconnected with other similar tools like ORCID profiles and how so, and also some national CRIS systems[...].” - Petra

Participants provided ideas for solutions. It was proposed that the interface be reversed, with users viewing the narrative CV element first. This is the description of one’s career in their own words. This prompted the group to focus on the discussion around narrative CVs, which can be enabled by the service. They did not object to the use of quantitative indicators, but only as a secondary channel of information.

¹³ Access the BIP! Scholar platform here: <https://bip.imsi.athenarc.gr/scholar>

¹⁴ Blue-sky thinking definition via Wiktionary: https://en.wiktionary.org/wiki/blue-sky_thinking - Blue-sky thinking refers to imaginative, unconstrained exploration of ideas without immediate regard for practical limitations or feasibility

“And if we invert the process, and we don’t start from data, but we start from the narrative, and then we use data to complement, then I think that this could be a very useful tool for humanities scholars as well.” - Chiara

Regarding the content that could be included on the platform for monitoring and potentially creating indicators, participants provided input on what would be positively received by SSH scholars. They suggested ideas related to book publishing, a key component of SSH, such as the platform showcasing derivatives of a published book, including the number of translations or a list of book reviews. Other suggestions focused on participation in projects, reports, and open reports from peer review. In terms of reviewing, it is viewed as a more qualitative form of assessment. There was a clear shift away from simplistic quantitative measures towards a more comprehensive evaluation framework that includes diverse outputs and qualitative assessments.

4.5 Open Science dimension of outputs and indicators

This theme was primarily addressed in CW3, where we deliberately focused on Open Science as a topic of discussion. The workshop developed along three axes: 1) which outputs could be considered friendly towards Open Science, 2) what is the role of narrative CVs in incorporating those outputs, and 3) what indicators could be created to measure those outputs. The discussion of Open Science outputs examined the evolving nature of research contributions and the importance of recognising a range of outputs beyond the traditional journal article format. Participants discussed the potential inclusion of non-traditional forms of research outputs, such as databases, preprints, and even multimedia formats like zines, which could facilitate the communication of research beyond the confines of academic circles.

“I strongly believe that publishing in national languages and in a way that is accessible and interesting outside academia, it’s very relevant to the work that we do, especially when we are being funded by European and national funding institutions, because we have an obligation to give back knowledge that we create to the wider society and not just academia. So when I think of Open Science, I think mostly of getting the results across outside the scientific community.” - Filipa

FAIR datasets were included in the discussion, as a key part of the commitment to Open Science, from the perspective of data sharing. One participant mentioned preprints and the ongoing debate about their integration into Open Science. Finally, the group moved the discussion to open peer review and the need for enhanced training and advocacy to address concerns about transparency and accountability and to alleviate fears regarding the impact of open peer review on individual careers.

Furthermore, the acknowledgement of contributions made by researchers beyond the traditional concept of authorship emerges as a pivotal theme. Some attention was given to extending the roles included in the CRediT¹⁵ taxonomy to encompass activities such as data curation,

translation and peer review, which are vital in reforming research assessment practices. The discussion on narrative CVs and Open Science has also identified a need for more qualitative and flexible tools in academic evaluation. It is often the case that traditional metrics, such as citation counts and journal impact factors, are inadequate for representing the breadth of research contributions in the SSH. The participant from Portugal noted that a narrative CV model¹⁶ has been successfully implemented for several years, enabling academics to document their career trajectories, research outputs and societal contributions in a bespoke manner. Participants also acknowledged that challenges remain in developing specific indicators to quantify these contributions.

“And contrary to Portugal, ... the narrative CV is not used in Greece. We are a bit old-fashioned. But in terms of the social sciences, I think that an interesting aspect also in relation to what we’ve put on the board is the interactions with society and NGOs. A lot of the work that social scientists are doing are about societal challenges, and they work closely either with NGOs or other actors outside academia. So I think that this is something that in the, let’s say, more traditional CVs are not that visible... I don’t know how we could translate this into specific indicators.” - Despina

The discussion progressed to potential indicators for assessing Open Science outputs in the SSH, with a particular emphasis on striking a balance between quantitative metrics and the need for qualitative assessments. They mentioned that even if we create indicators that measure Open Science outputs, there was concern that relying too heavily on quantitative measures, such as counting publications or datasets, could result in superficial assessments and encourage behaviours such as producing data simply to meet targets. The value of integrating qualitative aspects, such as the tangible impact, quality and usability of research outputs, was consistently highlighted.

There was a consensus that research assessment systems should be more sophisticated and flexible, taking into account the diverse research practices in SSH. The discussion emphasised the need for an approach that goes beyond simply counting outputs, focusing instead on how research contributes to broader impacts, such as influencing policy or being reused by others. This qualitative aspect was seen as crucial for fair and accurate assessments.

5. Discussion and conclusions

5.1. Key points of discussion

Several key issues emerged from the workshops. We note the need to define the type of evaluation being conducted, such as individual, project-based, institutional or other. It was important for participants to have a clearer context for the discussion, in line with what the SCOPE framework proposes, as different criteria and indicators become relevant at different points in the assessment spectrum. When participants were asked to discuss indicators as statistical proxies (National Research Council 2012), they instead mentioned outputs or activities to be

included in an assessment exercise. As a result, discussions around indicators *per se* were superficial. This was further evidenced by the clear denunciation of quantitative indicators during the workshops. Given how indicators are seen as quantifiable tools, there was an obvious struggle to determine indicators that express quality. It appears that the word “indicator” in and of itself is related to a numerical value and that value cannot reflect quality. Assessing quality remains a sensitive issue (Galleron et al. 2017), and participants agreed it is hard to define quality as a concept. There was agreement as we noted in the previous point, that it is hard, if not impossible, to create quantitative indicators that can capture the quality of one’s work. Quality could be associated with reusability at times, but it was noted that while there is a lot of focus and emphasis on producing open science outputs, it is not guaranteed that those outputs would be of the highest quality. Unsurprisingly, peer review was mentioned constantly as a concept to be included in the processes of research assessment. The rationale for peer review is that it is a quality assurance measure. However, peer review has also faced criticism recently (Abramo 2024; Ioannidis and Maniadis 2023), specifically due to the advent of AI. To address the challenges posed by the advent of AI (Checco et al. 2021), it was suggested that peer review could evolve from its current state and prioritise open peer review. Open peer review is also an expression of the open science movement, and when considering the project’s objective, open peer review became a critical point of discussion, how it can become more established, and the challenges that come with it (Ross-Hellauer 2017). However, limitations and concerns regarding the scalability of peer review remain (Abramo 2024; Ioannidis and Maniadis 2023), as well as questions surrounding the feasibility of eliminating entirely quantitative representations of success.

Overall, it became evident that when individuals had the opportunity to discuss research assessment in their own terms, they identified several areas that require improvement. A substantial number of alternative activities and engagements within and outside of academia were proposed as worthy of consideration during the assessment process. They articulated a sense of fatigue regarding the current emphasis on measuring performance through the lens of purely quantitative indicators.

Regarding open peer review, it would be prudent to further explore the literature and conduct additional work to determine when and where it should be integrated into the evaluation process, and whether it could serve as a point at which quality criteria are agreed upon. A similar study in the field of psychology produced recommendations for a phased approach to assessing individuals for hiring or promotion (Schönbrodt et al. 2022).

Monitoring in itself is not a bad practice. The issue lies in how much emphasis evaluation panels place on this technique, what is monitored (Rafols 2024), and whether we have truly progressed since the inception of DORA. DORA initiated the discussions that publication venue should not be a criterion of quality. CoARA made explicit mention to the central role peer review should play in the assessment. CoARA also did not exclude bibliometrics and indicators altogether. Finding a balance between quantity and quality still seems elusive.

5.2 Towards research assessment criteria for the SSH: recommendations

From the discussion and the analysis, we present an initial formulation of recommendations for the generic research assessment criteria for the SSH produced as part of the project. The approach is layered, rather than progressive, and the list of recommendations is

open to critique. We believe our recommendations align with those of Ochsner, Hug, and Galleron (2017) and Galleron et al. (2017), the CoARA Commitments (2022), and the SCOPE Framework (Himanen et al. 2024).

Context-consideration: Considering the specific context of the research assessment is essential before conducting any research assessment exercise. The context determines the indicators and methods of assessment. We propose that interested departments organise their own SCOPE or HumetricsHSS workshop as a starting point to produce an appropriate values-based framework for research evaluation.

Recommendations¹⁷:

1 - Indicators alone should not be used to determine the quality of a research activity. Quantitative indicators can be used in an evaluation process, but their weighting must be clearly defined and balanced with peer review.

2 - Peer review is an essential part of the assessment process. In any evaluation process, consider when there will be intervention by an expert panel to assess the quality of the individual's work.

3 - Narrative CVs should be explored and offered as an alternative to traditional CVs. Narrative CVs offer the opportunity to scholars to describe their contributions in ways that traditional CVs cannot do. As identified earlier, the quality assurance process of any exercise should be the peer review process, in addition to quantitative evidence.

4 - Expand the list of outputs, processes and activities that could be included in assessment exercises. We understand this requires a re-evaluation of proceedings and how assessment happens, but it has been noted that there are diverse activities that ought to be measured. By establishing criteria for inclusion of diverse processes and activities, it will bring into scope a lot of work done by SSH Researchers. It will also create a level-playing field with consistent rules for the assessment exercise to be considered transparent, ethical and equitable.

5.3 Concluding remarks

This paper serves as a reminder that even though topics related to SSH and research assessment have been discussed for two decades, they remain unresolved. The SSH domain plays an important role in this discussion due to different challenges of including diverse outputs, considering processes, and understanding how to assess the impact, excellence and quality of research. As a result, we have a more diverse perspective on context and practices. While we recognize challenges with established bibliographic databases, it is also important to consider alternative approaches to perform research assessment that reduce the focus on quantitative measures. This paper provides a clear documentation of the voices of those involved, respecting one of the three key principles of SCOPE: evaluate with the evaluated. Often research provides more questions than answers. Appropriately, we conclude this discussion with a quote from one of the participants, reflecting their sentiment about the challenge ahead in relation to research assessment: *“Do we need a revolution or is a reform possible, regarding evaluation?”*

¹⁷ Further work under the SSH Pilot in GraspOs on recommendations and guidelines for Research Assessment can be found here: <https://doi.org/10.5281/zenodo.17951789>

References

- Abramo, Giovanni. 2024. "The forced battle between peer-review and scientometric research assessment: Why the CoARA initiative is unsound." *Research evaluation*, May. <https://doi.org/10.1093/reseval/rvae021>.
- Agate, Nicky, Christopher P. Long, Bonnie Russell, Rebecca Kennison, Penelope Weber, Simone Sacchi, Jason Rhody, and Bonnie Thornton Dill. 2022. "Walking the talk: Toward a values- aligned academy." *Knowledge commons*. <https://hcommons.org/deposits/objects/hc:44632/datastreams/CONTENT/content>.
- Aksnes, D. W., Langfeldt, L., & Wouters, P. 2019. Citations, citation indicators, and research quality: an overview of basic concepts and theories. *Sage Open*, 9(1). <https://doi.org/10.1177/2158244019829575>.
- Archambault, Éric, and Étienne Vignola Gagné. 2004. "The use of bibliometrics in the social sciences and humanities: final report." *Science-Metrix*. https://www.science-metrix.com/pdf/SM_2004_008_SSHRC_Bibliometrics_Social_Science.pdf.
- Archambault, Éric, Étienne Vignola-Gagné, Grégoire Côté, Vincent Larivière, and Yves Gingras. 2006. "Benchmarking scientific output in the social sciences and humanities: the limits of existing databases." *Scientometrics* 68 (3): 329–42. <https://doi.org/10.1007/s11192-006-0115-z>.
- Australian Research Council. 2019. "Non-traditional research outputs (NTRs)". *State of Australian University Research*. <https://dataportal.arc.gov.au/era/nationalreport/2018/pages/section1/non-traditional-research-outputs-ntros/>.
- Balboa, Luciana, Elizabeth Gadd, Janne Pölönen, Karen Stroobants, and Erzsébet Tóth-Czifra. 2024. "The role of scientometrics in the pursuit of responsible research assessment." *LSE Impact Blog*. September 4, 2024. <https://blogs.lse.ac.uk/impactofsocialsciences/2024/09/04/the-role-of-scientometrics-in-the-pursuit-of-responsible-research-assessment/>.
- Bingham, Andrea J. 2023. "From data management to actionable findings: a five-phase process of qualitative data analysis." *International journal of qualitative methods* 22 (22).
- Checco, Alessandro, Lorenzo Bracciale, Pierpaolo Loreti, Stephen Pinfield, and Giuseppe Bianchi. 2021. "AI-assisted peer review." *Humanities and social sciences communications* 8 (1): 1–11. <https://doi.org/10.1057/s41599-020-00703-8>.
- CoARA. 2022. "Agreement on reforming research assessment." CoARA. July 20, 2022. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf.

- Curry, Stephen, Sarah de Rijcke, Anna Hatch, Dorsamy (Gansen) Pillay, Inge van der Weijden, and James Wilsdon. 2020. "The changing role of funders in responsible research assessment: progress, obstacles and the way ahead." Rori.figshare.com. November 18, 2020. <https://doi.org/10.6084/m9.figshare.13227914.v2>.
- DORA. 2012. "San Francisco Declaration on Research Assessment." <https://sfdora.org/read/>.
- Engels, Tim C. E., Tryuken L. B. Ossenblok, and Eric H. J. Spruyt. 2012. "Changing publication patterns in the social sciences and humanities, 2000–2009." *Scientometrics* 93 (2): 373–90. <https://doi.org/10.1007/s11192-012-0680-2>.
- European Commission, European Research Executive Agency. 2024. "Report on research assessment - Publications Office of the EU." Publications Office of the EU. 2024. <https://op.europa.eu/en/publication-detail/-/publication/219aa5ea-fae2-11ee-a251-01aa75ed71a1/language-en>.
- Galleron, Ioana, Michael Ochsner, Jack Spaapen, and Geoffrey Williams. 2017. "Valorizing SSH research: towards a new approach to evaluate SSH research' Value for Society." *Fteval Journal for research and technology policy evaluation* 44 (September). <https://doi.org/10.22163/fteval.2017.274>.
- Hicks, Diana. 2005. "The four literatures of social sciences." *Journal of management and social sciences* 1 (1). <https://ibtjbs.ilmuniversity.edu.pk/journal/jbs/1.1/1.%20The%20Four%20Literatures%20of%20Social%20Sciences.pdf>.
- Himanen, Laura. 2023. "GraspOS Deliverable 5.1 'Report on pilot setup, current practices and initial requirements.'" *Zenodo*, August. <https://doi.org/10.5281/zenodo.11091718>.
- Himanen, Laura, and Iiris Liinamaa. 2024. "GraspOS Deliverable D5.2: Pilot Findings and Progress Report." *Zenodo*, August. <https://doi.org/10.5281/zenodo.13629147>.
- Himanen, Laura, Erica Conte, Marianne Gauffriau, Tanja Strøm, Baron Wolf, and Elizabeth Gadd. 2024. "The SCOPE framework – implementing ideals of responsible research assessment." *F1000Research* 12 (May): 1241–41. <https://doi.org/10.12688/f1000research.140810.2>.
- Information Commissioners Office. 2021. "Introduction to anonymisation draft anonymisation, pseudonymisation and privacy enhancing technologies guidance." <https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf>.
- Ioannidis, John P. A., and Zacharias Maniadis. 2023. "In defense of quantitative metrics in researcher assessments." *PLOS Biology* 21 (12): e3002408–8. <https://doi.org/10.1371/journal.pbio.3002408>.

- Joachim Schöpfel, and Otmane Azeroual. 2024. "Responsible research assessment and research information management systems." *Encyclopedia* 4 (2): 915–22. <https://doi.org/10.3390/encyclopedia4020059>.
- Krüegel, Sybil. 2019. The informed consent as legal and ethical basis of research data production. FORS Guide No. 05, Version 1.0. Lausanne: *Swiss Centre of Expertise in the Social Sciences FORS*. doi:10.24449/FG-2019-00005.
- Leeuwen, van Thed. 2013. "Bibliometric research evaluations, Web of Science and the social sciences and humanities: a problematic relationship?" *Bibliometrie Scientometrics* 2 (September): 1–18. <https://doi.org/10.5283/bpf.173>.
- Mystakopoulos, Fotis. 2024. "The role of open science in fostering a more inclusive research assessment process in social sciences and humanities." *Zenodo*, January. <https://doi.org/10.5281/zenodo.10562683>.
- National Research Council. 2012. *Improving measures of science, technology, and innovation*. National Academies Press EBooks. <https://doi.org/10.17226/13358>.
- Ochsner, Michael, Sven Hug, and Ioana Galleron. 2017. "The future of research assessment in the humanities: bottom-up assessment procedures." *Palgrave communications* 3 (1). <https://doi.org/10.1057/palcomms.2017.20>.
- Pinto da Costa, Mariana. 2021. "Conducting cross-cultural, multi-lingual and multi-country focus groups: guidance for researchers." *International journal of qualitative methods* 20 (January): 160940692110499. <https://doi.org/10.1177/16094069211049929>.
- Rafols, Ismael. 2024. "The UNESCO Open Science Outlook: OS progresses, but unequally." [Leidenmadtrics.nl](https://www.leidenmadtrics.nl). February 2024. <https://www.leidenmadtrics.nl/articles/the-unesco-open-science-outlook-there-is-progress-in-os-but-it-is-unequal>.
- Reale, Emanuela, Dragana Avramov, Kubra Canhial, Claire Donovan, Ramon Flecha, Poul Holm, Charles Larkin, et al. 2017. "A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research." *Research evaluation* 27 (4): 298–308. <https://doi.org/10.1093/reseval/rvx025>.
- Rijcke, Sarah de, Clemencia Cosentino, Robin Crewe, Carlo D'Ippoliti, Shaheen Motala-Timol, Noorsaadah Binti A Rahman, Laura Rovelli, David Vaux, and Yao Yuy-peng . 2023. "The future of research evaluation: a synthesis of current debates and developments," May. <https://doi.org/10.24948/2023.06>.
- Rodriguez, Katrina L., Jana L. Schwartz, Maria K. E. Lahman, and Monica R. Geist. 2011. "Culturally responsive focus groups: reframing the research experience to focus on participants." *International journal of qualitative methods* 10 (4): 400–417. <https://doi.org/10.1177/160940691101000407>.

- Ross-Hellauer, Tony. 2017. "What is open peer review? A systematic review." *F1000Research* 6 (588): 588. <https://doi.org/10.12688/f1000research.11369.2>.
- Schmoch, Ulrich, Torben Schubert, Dorothea Jansen, Richard Heidler, and Regina von Görtz. 2010. "How to use indicators to measure scientific performance: a balanced approach." *Research evaluation* 19 (1): 2–18. <https://doi.org/10.3152/095820210x492477>.
- Schönbrodt, Felix D, Anne Gärtner, Maximilian Frank, Mario Gollwitzer, Malika Ihle, Dorothee Mischkowski, Le Vy Phan, et al. 2022. "Responsible research assessment I: implementing DORA for hiring and promotion in psychology." *psyArXiv (OSF Preprints)*, November. <https://doi.org/10.31234/osf.io/rgh5b>.
- Sharma, Mohit, Anurag Sarin, Priyanka Gupta, Shobhit Sachdeva, and AnkurV Desai. 2014. "Journal impact factor: its use, significance and limitations." *World journal of nuclear medicine* 13 (2): 146. <https://doi.org/10.4103/1450-1147.139151>.
- Sivertsen, Gunnar. 2016. "Patterns of internationalization and criteria for research assessment in the social sciences and humanities." *Scientometrics* 107 (2): 357–68. <https://doi.org/10.1007/s11192-016-1845-1>.
- Snijder, Ronald, 2019. "The deliverance of open access books: examining usage and dissemination". https://doi.org/10.26530/oapen_1004809.
- Sorbonne University. 2024. "Sorbonne University unsubscribes from the Web of Science." Sorbonne Université. 2024. <https://www.sorbonne-universite.fr/en/news/sorbonne-university-unsubscribes-web-science>.
- Stam, Alexandra, and Pablo Diaz. 2023. "Qualitative data anonymisation: theoretical and practical considerations for anonymising interview transcripts." *FORS Guide No. 20 Version 1.0 (March)*. https://forscenter.ch/wp-content/uploads/2023/03/qualitative-data-anonymisation_final.pdf.
- Tatum, Clifford, Zeynep Anli, Ludo Waltman, Anna-Kaisa Hyrkkänen, Janne Pölönen, and Josephine Nordling. 2023. "GraspOS Deliverable 2.2: Open Science Assessment Framework (OSAF)." *Zenodo*, December. <https://doi.org/10.5281/zenodo.11091512>.
- UNESCO. 2023. "Open Science Outlook 1: status and trends around the world." *UNESDOC Digital Library*. <https://doi.org/10.54677/giic6829>.
- Vergoulis, Thanasis, Serafeim Chatzopoulos, Giulia Malaguarnera, Zenia Xenou, Leonidas Pispiringas, Harry Dimitropoulos, Silvio Peroni, et al. 2024. "GraspOS Deliverable D3.2: Tools and Services." *Zenodo*, August. <https://doi.org/10.5281/zenodo.13619176>.

- Vergoulis, Thanasis, Serafeim Chatzopoulos, Kleanthis Vichos, Ilias Kanellos, Andrea Mannocci, Natalia Manola, and Paolo Manghi. 2022. "BIP! SCHOLAR: a service to facilitate fair researcher assessment." In *JCDL '22: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*. <https://doi.org/10.1145/3529372.3533296>.
- Wang, Ruishu. 2023. "Research on problems of scientific research performance evaluation mechanism of humanities and social sciences." *Frontiers in Educational Research* 6 (18). <https://doi.org/10.25236/fer.2023.061823>.
- White, Howard D., Sebastian K. Boell, Hairong Yu, Mari Davis, Concepción S. Wilson, and Fletcher T.H. Cole. 2009. "Libcitations: a measure for comparative assessment of book publications in the humanities and social sciences." *Journal of the American Society for Information Science and Technology* 60 (6): 1083–96. <https://doi.org/10.1002/asi.21045>.

Sažetak

Prema reformi vrednovanja znanstvenog rada u društvenim i humanističkim znanostima: napredak i uvidi iz projekta GraspOS

Cilj. Ova studija usmjerena je na razvoj kriterija za vrednovanje znanstvenog rada u društvenim i humanističkim znanostima (SSH) kroz projekt GraspOS. Cilj je izmijeniti tradicionalne metode vrednovanja koje se uvelike oslanjaju na bibliometrijske pokazatelje i izraditi preporuke prilagođene SSH zajednici, s posebnim osvrtom na ograničenja trenutnih praksi koje pretežno naglašavaju kvantitativne metrike.

Pristup/metodologija/dizajn. Ova istraživanje predstavlja rezultate dobivene kroz tri radionice održane s ispitanicima iz različitih SSH područja. Radionice su omogućile vođene rasprave o vrednovanju temeljenom na vrijednostima, raznolikosti rezultata i uključivanju kvalitativnih pokazatelja u vrednovanje. Prikupljeni podaci analizirani su deduktivnom metodom, s unaprijed definiranim kodovima za analizu.

Rezultati. Analiza je utvrdila usuglašeni stav ispitanika o potrebi za izradom uključivijeg okvira za vrednovanje znanstvenog rada koji prepoznaje različite rezultate, pored tradicionalnih članaka u časopisima. Ispitanici su naglasili važnost kvalitativnih mjera, poput recenzije, i predložili korištenje narativnih životopisa kako bi se bolje utvrdili pojedinačni doprinosi.

Ograničenja istraživanja. Istraživanje je provedeno unutar specifičnog konteksta pokreta za odgovorno vrednovanje znanstvenog rada (RRA) i projekta GraspOS, koji možda ne predstavljaju u potpunosti cijeli krajolik SSH. Generalizacija rezultata i zaključaka na sve SSH discipline nije moguća.

Praktična primjena. Rezultati provedene studije oblikovani su u vidu preporuka za pripremu općih kriterija za vrednovanje znanstvenog rada kako su predložili znanstvenici s područja društvenih i humanističkih znanosti koji su sudjelovali u provedbi istraživanja.

Originalnost/Vrijednost. Ovaj rad pruža vrijedne uvide u razvoj pravednijih i učinkovitijih praksi vrednovanja u akademskoj zajednici za širu zajednicu društvenih i humanističkih znanosti.

KLJUČNE RIJEČI: društvene i humanističke znanosti, istraživačka kultura, odgovorno vrednovanje znanstvenog rada, otvorena znanost