

Appearing as: Ad Prins, Jack Spaapen, Thed van Leeuwen, Nelleke van den Broek-Honingh, QRiH: Towards a fitting system for humanities research evaluation, ftEval Journal for Research and Technology Policy Evaluation, Issue 48, July 2019, p. 91-97

QRiH: Towards a fitting system for humanities research evaluation¹

Ad Prins*, Jack Spaapen**, Thed van Leeuwen***, Nelleke van den Broek-Honingh****

* info@adprins.nl, Support in ResearchManagement, Oosterkade 8b, 9711 RS Groningen, The Netherlands

** jack.spaapen@knav.nl, Royal Netherlands Academy of Arts and Sciences, Kloveniersburgwal 29, 1011 JV Amsterdam, The Netherlands

*** leeuwen@cwts.leidenuniv.nl, Centre for Science and Technology Studies Leiden University, P.O. Box 905, 2300 AX Leiden, The Netherlands

**** n.vandenbroekhoningh@rathenau.nl, Rathenau Institute, Anna van Saksenlaan 51, 2593 HW The Hague, The Netherlands

I. Introduction

In this paper we report the design and introduction of a new system for the assessment of Quality and Relevance in the humanities (QRiH) in the Dutch evaluation context and report also the first experiences of using the system in ongoing evaluations. The design of the QRiH system is an attempt to meet two challenges: The first is to find an evaluation method that fits the ways in which humanities researchers communicate with science and society. In many of the current evaluation systems, with the usual attention to metrics fitting the characteristics of research in Science, Technology, Engineering and Mathematics (STEM) fields, the research and communication practices in the humanities are hardly acknowledged. The second challenge is to deal with the increasing pressure on researchers, in humanities and all other fields, to demonstrate the societal relevance of research, while at the same time there is a lack of consensus on how to assess research with regard to the societal mandate towards greater attention for knowledge utilization. We describe specific characteristics of research communication in humanities and address how the communities of researchers and policy makers have been involved in a bottom up development. Also, we will argue that the format of the narrative for self-evaluation addresses the above challenges and gives room to the diversity in the communication outcomes among the research units. The first experiences in using the system are encouraging, but demand sustained attention of panels, researchers and policy makers in making the system valid.

II. The context in which QRiH emerged

The report *Duurzame Geesteswetenschappen (Sustainable humanities)* of the Committee on the “National Plan for the future of the Humanities”, also known as the Committee Cohen (*Committee on the National Plan for the Future of the Humanities 2009*), observed that, in terms of research assessment, the humanities are too much at the mercy of models derived from the exact sciences and medicine. The Committee recommended that the humanities develop its own set of assessment standards. In the years that followed, the Royal Netherlands Academy

¹ The authors thank Frank van Vree, chair of the working group, for his energetic and forceful efforts in developing QRiH and in realising the conditions for the working group to do so.

of Arts and Sciences KNAW took this task upon itself and tackled this assignment via several studies that culminated in a proposal for an adequate evaluation system for humanities research in 2012 (*Quality and relevance in the Humanities 2012*). This proposal has been quite influential, as it included as one of the first national systems both the academic and the societal dimension of scholarly activity. Two other sections in the Academy, the fields of design and engineering and the social sciences, developed similar visions on research at the same time. This work was also reported (KNAW 2010 and KNAW 2013) and together these fields influenced to a great extent the model that was adopted in the new Standard Evaluation Protocol 2015-2021 (SEP), launched in 2014.

The SEP 2015-2021 enables a balanced assessment of both the academic quality and the relevance to society. These two are the main criteria for the evaluation by an international review committee, which is based on a self-assessment report. SEP Table D1 (table 1) forms the basis for providing evidence in support of the self-assessment report. The indicators in this table determine the content of the self-assessment report and the information gathered for the report forms the basis for the evaluation by the assessment committee.

Table 1. Quality domains and assessment dimensions of the Dutch Standard Evaluation Protocol (SEP 2015 – 2021)

		Quality domains	
		Research quality	Relevance to society
Assessment dimensions	Demonstrable products	1. Research products for peers	4. Research products for societal target groups
	Demonstrable use of products	2. Use of research products by peers	5. Use of research products by societal target groups
	Demonstrable marks of recognition	3. Marks of recognition from peers	6. Marks of recognition by societal target groups

The indicators given in the table are given as yet empty categories, which can be used in a field-dependent fashion. Each discipline can bring up indicators that suit best production and communication practices of the field.

The KNAW reports, and the subsequent SEP protocol, were published in a wider context of increasing criticism – such as the San Francisco Declaration On Research Assessment (DORA), the Leiden Manifesto, and the Science in Transition movement – on the perverse influences of research metrics in the science system, among other the ever-increasing drive to “publish or perish”. As a consequence of this criticism, productivity which used to be a main quality criterion, now has been taken out of the SEP 2015-2021, while societal relevance gained an equal status to scientific quality in the evaluation model.

One of the main characteristics of the SEP is that it presents a general framework, but within this framework leaves room for the various disciplines to develop a set of criteria and indicators that fit best the production and communication habits of the field.

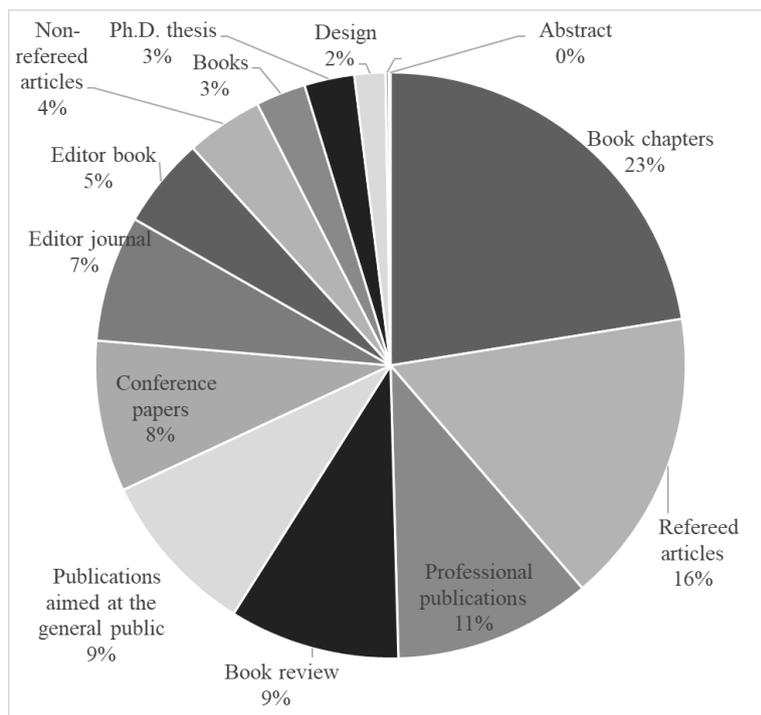
In this context a small working group was assigned by the humanities deans to develop a set of assessment standards specifically for the humanities.¹ The working group worked along three lines of activity. First, we researched production and communication practices in the Dutch academic humanities research. Second, we looked at what was happening in a few neighboring

countries, Norway and Flanders (Current Research Information System in Norway – CRISTiN and VABB database systems), and the United Kingdom (focusing on the impact pathways developed in the British Research Excellence Framework – REF 2014). Third, we developed QRiH in a way that this would fit into the current national evaluation system for academic research, SEP 2015-2021. And we did this bottom-up, that is we engaged the research and policy community in the humanities. The three lines of activity will be explained in more detail in the following paragraphs.

III. Research production and communication in the humanities

A brief analysis of the production and communication practices of two large faculties of humanities (Leiden and Amsterdam) shows a wide diversity in types of output and use of languages, and also indicates difficulties when relying on resources that are often used in the evaluation of STEM fields, such as Web of Science. The research information systems of these faculties show – in line with other research (Van Leeuwen 2013) – that the largest portion of research output is not in journals but in book chapters, and lists a wide variety of other types of research outcomes, such as books, professional publications, book reviews, or publications aimed at the general public (figure 1).

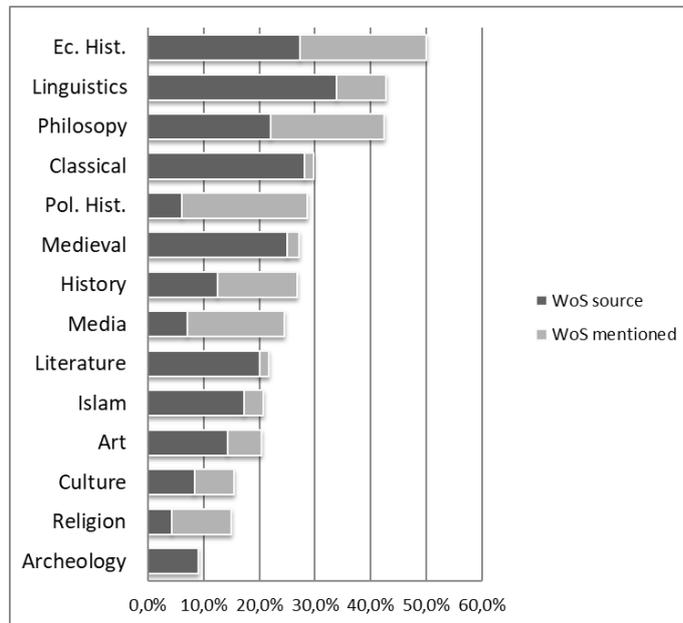
Figure 1. Relative sizes of research output per type, Leiden and Amsterdam Universities



Also, the output in peer reviewed journals, accounting for 16% of the total of research output of the two humanities faculties, appears to be represented only to some degree in journals mentioned or listed in Web of Science (WoS). The share of Web of Science source journals and of journals mentioned in Web of Science also varies according to the domains of scholarly research. In Economic History, about half of the output in reviewed journals is in WoS journals,

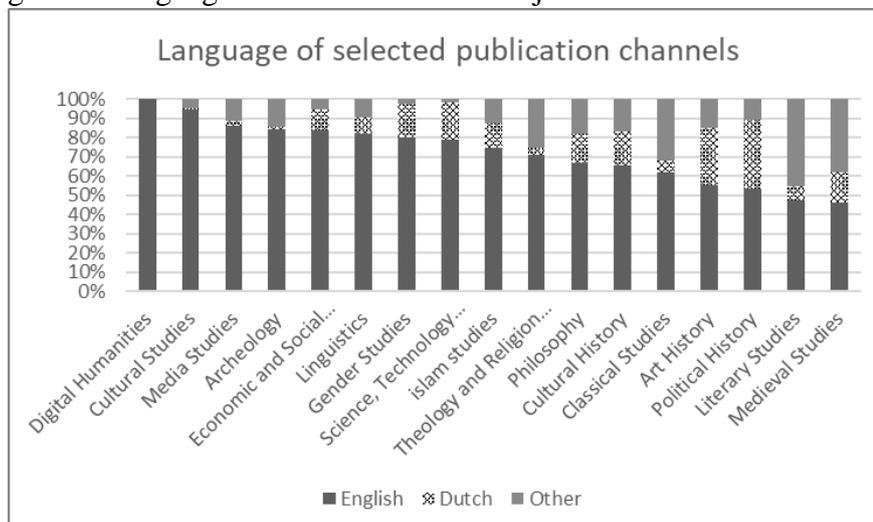
whereas in many other fields, such as Culture Studies, Religion and Theology Studies or among researchers from Archaeology the share of WoS journals is below 20%. (See figure 2.)

Figure 2. Share of output in journals with WoS source status or mentioned in WoS in Leiden and Amsterdam output. Total of papers in peer reviewed journals (100%).



Another well-known characteristic of humanities research is the fact that communication occurs in many more languages than English. This characteristic varies across its various domains. Considerable differences occur in the language of journals selected by panels in the field of humanities researchers as prominent publication channels (see also paragraph IV). For example, the selection for digital humanities consists almost completely of journals in the English language oriented, while Medieval studies shows that more than half of the selected journals is in other languages or in Dutch (see figure 3).

Figure 3. Language orientation of selected journals in 17 humanities fields



These findings clearly indicate that classical bibliometric approaches such as citation analysis or WoS status of journals are insufficient to support research assessment in the humanities, which is consistent with other literature: a similar study of different publication cultures among Social Sciences and Humanities (SSH) and STEM researchers, based upon bibliometric analysis of reference behavior by authors publishing in journals covered in the WoS, shows the lesser relevance of journal based assessment in SSH compared to STEM: in SSH, between 10% and 40% of all references are addressed to journal literature in WoS, while for STEM this amounts up to 95% (van Leeuwen 2013). Additionally, taking it from the perspective of output produced by a whole university, classical bibliometric analysis based upon WoS makes the research conducted in most SSH departments nearly invisible, while the internal output registration system clearly shows the presence of a wide variety of scholarly communication types being present (van Leeuwen et al. 2016). This situation disqualifies the existing bibliometric toolbox for SSH and law research assessments, as quantitative analysis only deals with a very small portion of what actually has been produced, across a variety of communication channels.

One of the assumptions at the start of the QRiH project was that in the humanities, publications for wider audiences and for students can be regarded as a very important expression of societal relevance, even to the point that the line between academic and non-academic publications often is difficult to draw (Sivertsen 2016). The working group took this idea further by proposing the category of “hybrid” publication as a relevant category for humanities, defined as publications with scholarly status also addressing wider audiences of academics and non-academic readers.

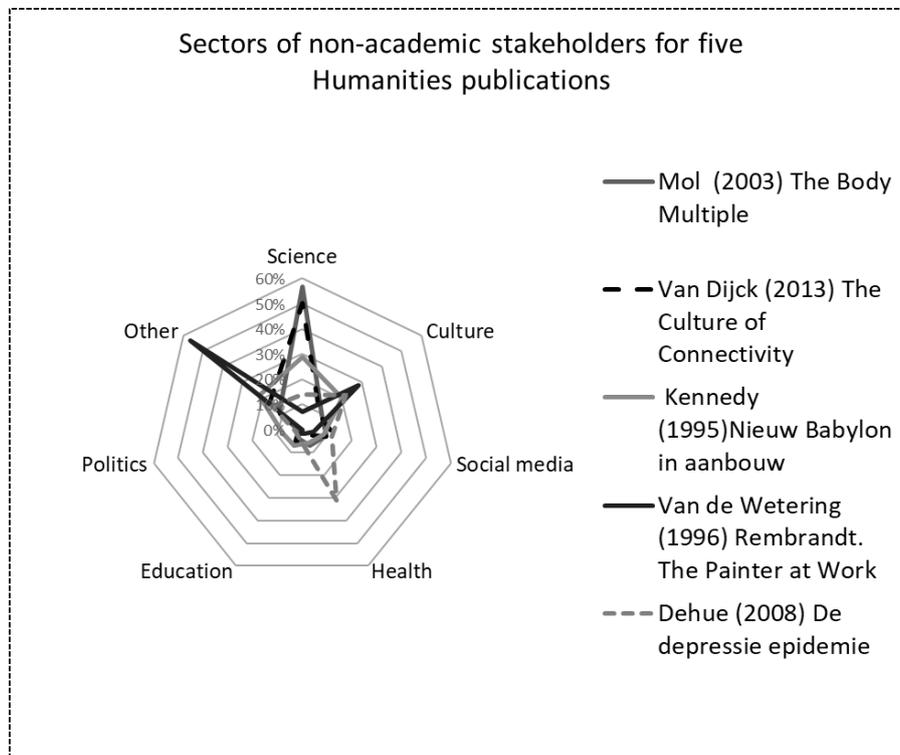
The claim of the hybrid characteristics of humanities publications can be substantiated. In a questionnaire send out to panels in the field of humanities researchers (see also paragraph V), the various participating panels reacted positively to the request to identify examples of such “hybrids”. Several of the suggested works were subsequently analysed for references in scholarly literature (using Google Scholar) and references to be found in non-academic environments using the search engines Google and BING (Prins et al. 2016), demonstrating the actual use in both the scientific and societal sphere.

Table 2. Five frequently used humanities publications used both on internet and cited by Google Scholar.

	Google Scholar cites	# net societal stakeholders*
Annemarie Mol (2003) <i>The Body Multiple</i> Duke University Press	3359	132
José van Dijck. <i>The Culture of Connectivity: A Critical History of Social Media</i> . Oxford: Oxford UP, 2013.	729	132
James C. Kennedy, <i>Nieuw Babylon in aanbouw. Nederland in de jaren zestig</i> (Amsterdam 1995: Boom)	280	153
Ernst van de Wetering. <i>Rembrandt. The Painter at Work</i> , AUP, 1996.	150	170
Trudy Dehue (2008) <i>De depressie epidemie</i> , Amsterdam: Augustus	103	215

* *Net societal stakeholders: Libraries, repositories, web shops and other internet finds not-relevant for meaningful communication are excluded from these results. Also excluded are references from scholarly journals.*ⁱⁱ

Figure 3. Use of humanities publications by non-academic stakeholders according to their sector. 100% = total of use by relevant stakeholders



These examples illustrate the need to develop specific indicators for humanities research. The division between the academic and societal sphere is seemingly less clear or strict in the humanities, meaning that societal production in the humanities is not a spinoff derived from academic production, but can be an intricate outcome of scholarly production. The examples also show the mutual entwinement of academic and societal productivity in the humanities.

IV. Learning from other research evaluation systems

Developing the QRiH system also included reviewing comparable systems in some other European countries, such as the impact case studies in the British REF exercise, the publication databases Current Research Information System in Norway CRISTiN (used in Norway) and the Belgian Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen (VABB-SHW) used in Flanders. A comparison between the content of the databases and the ways in which publication and other output data are used in the allocation of funds is useful, even though, unlike the SEP, the indicators derived from the VABB-SHW and CRISTiN are applied in a performance-based research funding system (PRFS) distributing institutional grants to the universities (Ossenblok et al. 2012).

CRISTiN is the national research information system of Norway. It documents all scholarly articles by Norwegian researchers, and complements the library system database *BiBSYS*, which focuses on books. The Flemish VABB-SHW academic database is developed specifically for the social sciences and the humanities because it is felt that these fields were not adequately represented in WoS database, which serves as the basis for allocating funds

among STEM domains in Flanders. Both systems contain several thousands of journal titles and distinguish between them in different ways. The Norwegian system makes a difference between level 2 publications (in international journals) and level 1 publications (other journals, many of them Norwegian, that meet scientific criteria), and level 0 for non-scientific publications. Level 1 and 2 are indirectly tied to financial distribution in the universities. The Flemish system uses a similar distinction between WoS journals and non WoS journals. VABB-SHW is directly coupled to allocation of Bijzondere Onderzoeksfondsen (Special Research Fund, BOF) used to reallocate funding between the universities via points given to 5 different types of publications. Books get 4 points, articles 1. The policy context of these systems differs from the Dutch policy context the QRiH system has to operate in. But these systems are worthwhile investigating because they face partly the same problems QRiH faces. The main issue is how to value publication media that are not part of WoS or other international databases.

In the Dutch SEP evaluation system, a centralised database for journals and publishers or other bibliometric indicators is absent. Also, other than the British REF system (Sivertsen 2016), the evaluation outcomes do not include direct funding consequences between institutions. The implication of financial consequences of systems is that the information is very much focused on competitive elements and comparisons between groups of researchers which may be sensitive of the indicators used (Hammarsfelt et al. 2015). As Ossenblok et al. argued, researchers working in the Flemish VABB-SHW have published increasingly in WoS journals to the detriment of publications in the local language, following its rating system based on the WoS status. By contrast, the share of publications in Norwegian remained stable, occurring in the Norwegian CRISTiN systems that includes a stimulus to publish in the local language next to publishing in WoS journals (Ossenblok et al. 2012).

The Flemish and Norwegian systems also differ from the field orientation of QRiH to address the specific needs of the humanities only, as the VABB-SHW system intends to inform decision making for funding in both the humanities and the social sciences, and the CRISTiN system addresses the entire national field of scientific and scholarly research (Sivertsen 2016). However, there are also similarities between the two: the organisation of domain panels to include specific expertise, a centralised supervising body or authority, and a central collection of information on publications. QRiH adopts the first two of these similarities between CRISTiN and VABB-SHW.

Another important evaluation system is the 2014 British Research Excellence Framework (REF UK). In REF, experience has been gained of what are called *impact case studies*. These studies focus on the impact on society and describe, among other things, the project, the participants and their share in the project, the nature and scope of the impact, and what the project actually yields. The case study reports followed a specific structure and were no longer than 5 pages²

Looking at the UK REF exercise, we focused on these impact case studies that were introduced specifically to evaluate the societal impact of research. Impact was defined rather broadly as research having “*an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia*”³. The REF website has a database with about 7000 impact case studies, all written in the mandatory format. These functioned as a support to specifically underline the societal relevance of the research conducted.

² <http://impact.ref.ac.uk/CaseStudies/>.

³ <https://re.ukri.org/research/ref-impact/>

Comparison with these other systems led us to believe that some elements could be very useful in the Dutch situation while other were less suitable. For example, the fact that there is no direct connection between output and funding in the Dutch system we were able to refrain from levels for journals and other kinds of publications. Consequently, we were able to develop lists of journals and publisher without levels and without a metrics system. At the same time, we tried to be selective here and limit the number of entries in the QRIH database, now including 2210 unique entries for the most important journals and publishers. The selection was given to panels, an element we took from the Flemish and Norwegian systems. Furthermore, we decided to develop indicators for quality and relevance of research bottom up, that is, the research community – via the research schools – was asked to come up with proposals.

V. Bottom up: the involvement of expertise from the various domains.

To develop a detailed view on the publication cultures in the different research domains in the humanities, we sent a request to the boards of the 17 research schools in which Dutch humanities is organised to assemble panels of researchers – junior as well as senior and prominent ones -. Over the period of two years, more than 200 researchers have participated in these panels. We have asked the panels to answer a questionnaire about various aspects of their publication culture, including the importance of peer review, typical forms of output, and to list journals and book publishers relevant for the various audiences in their domain, such as publication channels aiming at specialties, disciplinary and multidisciplinary audiences and also de wider context of hybrid audiences (combining both academic audiences and general readership).

Overall, panels have reacted positively to the request but express also concerns that lists of journals and book publishers might lead to the development of a metrics-based system. The question about outcomes of research and communication typical for the scholarly domain has led to listings of various forms of communication usually overlooked in output counting. The lists include catalogues for museum exhibitions, films and documentaries, designs and software programmes and other forms of output.

In the various stages of developing the QRIH system, we have held meetings with board members and policy makers of Humanities faculties and with the boards of the research schools. Although these meetings have resulted in positive reactions about the involvement of the panels, concerns remained not only over the possibility that QRIH in the end would lead to a metrics system but also that the outcomes of the work of the panels could be too restrictive in cases of multidisciplinary scholarship, or with respect to domains not covered by the panels. The comments raised during the meetings, and in the numerous talks and phone conversation with policy makers and researchers eventually led to the proposal that the system should be based on the *narrative* as the leading format for self-evaluations.

QRiH in a nutshell

The basic structure of QRiH is the format of the narrative for the self-evaluation of the research unit. The narrative should address both the scientific and the societal mission of the research and be supported by concrete evidence. Indicators in the six cells of table 1 should be elaborated in ways that fit the humanities. The working group decided to publicise the different elements on a website () and use the website as a work in progress. Researchers in the humanities can use the website in SEP evaluations, and at the same time share experiences and do suggestions for improvement. This is what is happening right now because many of the humanities faculties currently are going through a SEP evaluation.

The narrative, much in common with the format developed in the British REF system, should allow the institute, the group or the programme to indicate what the core of the research is, how it should position itself and which strategy is being pursued in order to achieve the objectives and share the research results with the academic world and society, as well as the success of those results. The intention is that claims of productivity, use and recognition put forward in the narrative should be substantiated with evidence that can be derived from indicators proposed by the various domain panels and authorised by a national panel (*authorised indicators*) or by put forward self-formulated (*reasoned*) evidence with the help of a broad list of indicator definitions.

As QRiH is to be used in the context of SEP evaluation, its format of the narrative implies a slight but important alteration of the SEP format. The original SEP format consists of 6 cells, in which the various topics of the self-evaluation are to be elaborated (see table 1). Although the SEP format leaves open which kinds of evidence is to be put forward, thus leaving room for variation for the diverse academic disciplines, the format can be read as making a categorical distinction between the spheres of academic research and society, distinguishing Research Quality from Societal Relevance. For many scholarly activities in the humanities, however, this is too restrictive. The narrative of the QRiH aims to address this issue by allowing via the narrative form to demonstrate quality and relevance both in the academic and societal sense either as separate spheres or combined.

The system is sustained by a website⁴ () providing information about authorised and reasoned indicators for each cell in the SEP protocol: there are indicators for products – for peers and for societal parties, indicators of use by peers and in society, and indicators for recognition in both spheres. A template for the narrative can be downloaded to structure the information in the narrative in such way that it will fit the size of maximum 15 pages required according to the SEP protocol. And the website provides various tools, including examples of hybrid publications, domain profiles and lists of journals and publishers selected by the domain panels as exemplary for the communication among specialties, the domain or in multidisciplinary communication.

VI. Intermediary review of QRiH

In 2018 most of the Dutch humanities research units are to be assessed according to the SEP 2015-2021. The draft version of QRiH was introduced in December 2017 and the working group agreed with the deans of the humanities faculties that research units to be assessed were

⁴ <https://www.qrih.nl/en>

going to use the QRiH as a guiding principle. This offered the opportunity to inquire after the first user experiences. For this purpose, we drafted a questionnaire with questions about the usefulness of QRiH in preparing the self-evaluation. Of the twenty units to be evaluated in 2018-2019, so far, seventeen have actually prepared self-evaluations, and possibly have worked with the QRiH system. Fourteen have responded.

The preliminary impressions based on the 14 received and completed questionnaires are that QRiH is overall appreciated as a tool that gives humanities researchers the opportunity to report in a way that is representative for their activities, especially via the narrative. QRiH appears to be widely known by directors and policy makers; only one policy officer (new at the position) was not familiar with QRiH. Most respondents (11) indicated to have used QRiH (more or less extensively) while writing the self-assessment report. Two respondents indicated that they had not used QRiH because they had started their self-assessment procedure before QRiH was publicised. Most respondents indicated to have used the format of a narrative, which was received with enthusiasm. In general, the set of (qualitative and quantitative) indicators is experienced as helpful.

Some respondents indicate that there are too many different groups of indicators and that indicators for societal productivity, use and recognition should be more specified, preferably also in an authorised form. Also, as QRiH is developed with the help of domain panels from the various research domains, the distinctions among research domains that are visible on the website seem to be confusing for users, especially for research units with a more multidisciplinary focus. It also appeared that the domain profiles developed as a soft benchmarking tool by the domain panels, appeared hardly to be used, because they were not seen as relevant. Arguably, this relates to the fact that many research units cover several research domains.

Remarkably, the lists of journals and publishers, organised per research domain, in the other systems a guiding element, seem to be hardly used by Dutch researchers. The reason for this is not yet clear. It could be due to the grouping of journals and publishers in domains (and research units cover sometimes several domains), but also because not all people find the lists user friendly. Another reason could be that in some domains there is ongoing discussion about the content of the lists.

Additionally, in the contacts with some directors it appears that the distinction of QRiH and the SEP protocol is not yet clear enough, leading some to revert to the SEP protocol. In part it is argued that the SEP protocol is an established format, while others also indicate that the information systems for research output are aligned to the SEP protocol.

Although the first reactions in general are quite positive – in particular with regard to the possibilities offered by the narrative - some specific aspects of QRiH are hardly used or in need of further elaboration. To get a better sense of why some aspects are hardly used, in-depth interviews will be held in the fall of 2018.

VII. Conclusions and discussion

The intention of the QRiH system is to offer an assessment system specifically designed for the humanities, adapted to the specific characteristics of the scholarly work in the humanities and developed with the help of its researchers and policy makers. The system works in the Dutch context of evaluation, meant to function within the broader scope of the SEP evaluation protocol. This SEP protocol is characterised by an equal attention for both the academic and societal aspects of production, use and recognition of research and lacks the linkage of financial consequences that is typical to other systems. Although the Dutch SEP provides a format flexible enough for a wide variety of disciplines, its application for assessing research units in

the humanities has lacunae that QRiH intends to mend. A brief analysis of the characteristics of production and communication of humanities research reveals not only that the types of communication are far more diverse than journal articles, books and book chapters, or that the communication includes various languages: the distinction between academic and societal communication is often not very relevant, leading to forms of communication distinctly different than in STEM fields such as hybrid publications. The fact that we aim at developing a special indicator for hybrid publications met with considerable enthusiasm in discussions with researchers at the University of Amsterdam. Therefore, QRiH offers the possibility to address the various aspects of quality and relevance both in the scholarly and the societal spheres in a flexible narrative form. The format of the narrative is supported by sets of authorised and reasoned indicators including also lists of prominent channels of communication among specialists, in disciplinary and in multidisciplinary settings.

Developing and implementing the QRiH system not simply the introduction of a set of indicators to be duly applied by policymakers, researchers and committee members. The development and introduction of QRiH took – and still takes – place in complex sets of contexts each posing constraints and possibilities. The first is the existing structure and demands of the protocol for the evaluation of research in the Netherlands, the Standard Evaluation Protocol – SEP. Another highly relevant context are the available evaluation systems and approaches in other countries, such as the British Research Excellence Framework (REF), the Flemish VABB-SHW system, or the Norwegian CRISTiN approach. Yet another is the organisation of the field in universities, faculties, institutes and research schools, and the information available at each level or organisation to sustain and support a specific research evaluation. A most crucial context consisted of the expectations and anxieties of researchers in the field and of the board members of faculties, institutes and schools. The introduction of QRiH and the idea of a narrative to demonstrate quality and relevance of humanities research could take place by accommodating to each of these, and by challenging these contexts.

The development of QRiH is a long-term process, for two main reasons. First, both researchers and policy makers should feel as the owners of the system, for which we aimed to have a bottom up process. Second, developing the various parts of the system, in particular the indicators, is a demanding endeavor. From the reactions via the questionnaire and in conversations, it is safe to conclude that we are half way now. QRiH has shown the possibilities to design a system for the evaluation of research in the humanities that does justice to the disciplinary diversity of the field, and to the diversity of its outcomes and ways of communication. Its main characteristic, the possibility to demonstrate the academic and societal quality and relevance of research programmes in a comprehensive way via the narrative, guided by a format and a broad set of well described indicators, authorised or other, is well received among researchers, boards and policymakers.

We have reason to assume that the characteristic of the narrative has contributed to a change in expectations among researchers. At the start of the project, the attempt to formulate indicators for quality and relevance was met with distrust and anxiety among some researchers. In view of the absence of shared views about how research in the humanities is to be publicly accountable, combined with the dispute over research indicators in many countries, this was understandable. The bottom up process proved very valuable: by exchanging experiences and information with researchers and policymakers, and during the various discussions distrust gave way to critical apprehension, but also a raising sense that the new system provides possibilities for the better. Needless to say, the process of development and introduction of QRiH is still going on. The next steps will be the analysis of the questionnaires that were sent

out to all participating research schools and look at the consequences for QRiH. Also, the set of indicators will be elaborated further to strengthen the supporting evidence for the narrative.

ⁱ The working group was chaired by prof. dr. Frank van Vree (the UvA, now director of NIOD-KNAW) and consisted of Thed van Leeuwen, Ad Prins, Jack Spaapen, Martin Boeckhout and David Duindam. The KIG Academic Advisory Committee consisted of Prof. Remieg Aerts (RU, UvA), Emeritus Prof. Geert Booij (Leiden), Prof. Martha Frederiks (UU, 2014-2015), Prof. Mary Kemperink (RUG), Prof. Anthonie Meijers TU/e, and Prof. Frank van Vree (chair; UvA, NIOD). Nelleke van den Broek-Honingh is member of the present working group.

ⁱⁱ The numbers differ from those in Prins 2016, after classifications of more websites

VIII. References

- Commissie Nationaal Plan Toekomst Geesteswetenschappen (Commissie Cohen). (2009). *Duurzame Geesteswetenschappen*, Amsterdam University Press, Amsterdam
- Committee on the National Plan for the Future of the Humanities (Committee Cohen). (2009). *Sustainable Humanities*, Amsterdam University Press, Amsterdam
- Declaration, D. O. R. A. (2012). The San Francisco Declaration on Research Assessment. Putting science into the assessment of research.
- Dijstelbloem, H., Huisman, F., Miedema, F., & Mijnhardt, W. (2014). Science in Transition Status Report: Debate, Progress and Recommendations.
- Hicks, D., Wouters, P., Waltman, L., Rijcke, S. D., and Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*. [doi:10.1038/520429a](https://doi.org/10.1038/520429a)
- Koninklijke Nederlandse Akademie van Wetenschappen. (2010). *Kwaliteitsbeoordeling in de ontwerpende en construerende disciplines*, Amsterdam. [Quality Evaluation in design and engineering disciplines]
- Koninklijke Nederlandse Akademie van Wetenschappen (2012). *Quality and Relevance in the Humanities*. Towards an adequate system for the evaluation of research. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Koninklijke Nederlandse Akademie van Wetenschappen (2013). *Naar een raamwerk voor kwaliteitsbeoordeling van sociaalwetenschappelijk onderzoek*, KNAW Amsterdam
- Ochsner M., Hug, S., Galleron, I., (2017). *The future of research assessment in the humanities: bottom-up assessment procedures*. Palgrave Communications. 3:17020 doi: 10.1057/palcomms.2017.20.
- Ossenblok, T. L., Engels, T. C., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science—a comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation*, 21(4), 280-290.
- Prins, A.A.M. & Spaapen J. B. (2017?). Serving Variegated Audiences: From Ranking Oriented Evaluation to Mission Oriented Evaluation, *fteval Journal for Research and Technology Policy Evaluation*, 44, September 2017, pp 42-49
- Quality and Relevance in the Humanities. (2017). QRiH, an instrument for describing, systematically, quality and relevance in humanities research in the Netherlands. <https://www.qrih.nl/en>
- Dijstelbloem, H., Huisman, F., Miedema, F. & Mijnhardt, W. (2013). Science in Transition. Why science does not work as it should, and what to do about it, position paper. <https://scienceintransition.nl/en/>
- Spaapen, J.B., & Prins, A. (2016). Contextual evaluation of multi-, inter-, and transdisciplinary research. In: Hubert, B. and Mathieu N. et al. (Eds) *Interdisciplinarités entre Natures et Sociétés*, Peter Lang, 2016, p. 273-290
- Van Leeuwen, T.N. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie - Praxis und Forschung*, 2013. 1-18 (<http://www.bibliometrie-pf.de/article/viewFile/173/215>)

Van Leeuwen, T.N., van Wijk, E., & Wouters, P.F. (2016). Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university, *Scientometrics*, 106 (1), 1-16

VSNU, KNAW & NWO. (2014). Standard Evaluation Protocol 2015–2021, protocol for research assessments in The Netherlands. Voorburg: VSNU, KNAW & NWO.