

Evaluating academic research at the crossroads of academic and societal demands

Nelleke van den Broek Honingh¹, Jack Spaapen,² Ad Prins³

1 n.vandenbroekhoningh@rathenau.nl

Rathenau Institute, Anna van Saksenlaan 51, 2593 HW The Hague

2 jack.spaapen@knaw.nl

Royal Netherlands Academy of Arts and Sciences, Kloveniersburgwal 29, 1011 JV Amsterdam

3 info@adprins.nl

Support in ResearchManagement, Oosterkade 8b 9711 RS Groningen

Conducting academic research in the 21st century is both demanding and exciting. Demanding because a broad variety of stakeholders (from policy, industry, and the wider society) is penetrating university life with rising expectations. And exciting because the questions that are put forward by this complex societal context can only be addressed in a productive way by collaboration between academic disciplines (interdisciplinarity) and or even wider collaborations between academics and actors outside academia. This means that academics have to rethink their research in terms of its relevance for those stakeholders. It also means that funding organisations (governments, institutions, research councils) have to rethink their research policies and their evaluation models, since traditionally these policies and models mainly focused on academic knowledge production and communication channels. In this article, we will review the national evaluation system that has been developed in the Netherlands against this widening context in which academics have to perform nowadays.

The Dutch evaluation system

In 2003 the Dutch Government introduced the Standard Evaluation Protocol (SEP) as the national evaluation protocol for publicly funded research at universities and research institutes. The SEP aimed at combining the evaluation of the quality and societal relevance into one comprehensive system of research evaluation. Arguably, the Netherlands was the first country to do this. Through the SEP all academic research was to be evaluated every six years, and after each six year cycle, the SEP would be reviewed by the main participants (universities, Royal Netherlands Academy of Arts and Sciences, Netherlands Organisation for Scientific Research) and adapted if necessary. Currently, the third version of the SEP is running. (VSNU, KNAW, NWO, 2014)

While the SEP has become part and parcel of the Dutch Academic system, for an important part thanks to the fact that SEP is operated by the institutions themselves, in an autonomous way independent of the government, some scientific areas criticized the system for being not flexible enough for their particular ways of producing knowledge and communicating about it with a variety of stakeholders, most notably the humanities (but also the social sciences and the engineering fields). (Van der Hoeven et al., 2010; Bensing et al., 2013; Algra et al 2013)

One of the results is the development of QRiH, a system for the assessment of Quality and Relevance in the Humanities. (Prins et al, 2019)

QRiH: combining societal and academic contexts

The deans of the humanities faculties decided to design an evaluation protocol that was based on the SEP, but at the same time catered more specifically to the knowledge production system of the humanities. This variant of the SEP was called Quality and Relevance in the Humanities (QRiH) and was published by the end of 2017 on a dedicated website, <https://www.qrih.nl/en>. While QRiH maintained the main SEP distinction of scientific quality and societal relevance as main evaluation domains, two changes are rather critical.

The first is that QRiH takes the narrative as the main structure of the self-evaluation. While in the SEP the narrative is reserved solely for societal relevance, QRiH asks researchers to review their work in a comprehensive narrative paying attention to both the scientific and the societal context. Second, for QRiH a wide range of indicators has been developed and made available on the website. Indicators are both qualitative and quantitative, and regard scientific quality and societal relevance, including hybrid indicators for knowledge products and outcomes that addresses academic and non-academic audiences at the same time. QRiH also reviews a range of outcomes that are specifically characteristic for research in the humanities, such as catalogues, films, exhibitions etc.

One of the indicators exemplifies the comprehensive characteristic of the narrative, in its attempt to combine the scientific with the societal context of research outcomes. “Hybrid publications” in the sense of QRiH – other than referring to an open access characteristic – gives the opportunity to enlist output addressing general as well as scholarly audiences. Examples of this type of output are well written books, or other scholarly output aiming to address broad audiences with intellectual interests. The effects of this type can be demonstrated as well, with the Contextual Response Analysis, a method to trace the reception of academic output via various sources, such as dedicated databases, Google Scholar and via internet searches. (Prins and Spaapen, 2017)

Given the wide variety in humanities research, not only in terms of disciplines or fields but also in the way relevance is expressed, QRiH has been developed bottom up, via panels of researchers operating in 17 different humanities domains. Initial questionnaires for domain panels intended to chart the diverse features of publication cultures of each of the humanities domains, paying ample attention to diversity also with regard to specialized, disciplinary and multidisciplinary characteristics. One of the tasks of these domain panels was to produce limited lists of journals and publishers that represent certain quality standards and are deemed important in their specific domains. The panels were also asked to come forward with examples of hybrid publications and to provide information about research outcomes in other forms than publications in terms of papers in journals or scholarly books. Subsequently, various meetings were staged with panels and representatives in the various humanities fields, involving the cooperation of over 200 scholars.

Although the bottom up procedure in developing QRiH ensures a rich system that is well grounded in humanities research practices, and leads also to increasing approval among researchers, the proof of such a system is in its actual use. A first opportunity to review the experiences of users occurred at the end of 2018.

First experiences in using QRiH

Most humanities research units in the Netherlands have been evaluated in or around the year 2018. While the use of QRiH was not mandatory, we were curious to see whether and how humanities research units made use of QRiH writing a self-evaluation. To determine whether QRiH was used when writing the self-evaluations and to determine whether QRiH was perceived as helpful, we decided to evaluate the use of QRiH via a short questionnaire. The goal of the evaluation was threefold. The first goal was to determine whether (parts of) QRiH were used when writing a self-evaluation. Secondly, we wanted to know if QRiH was perceived by users as helpful when writing a self-evaluation. The third goal was to determine which topics presented in QRiH were perceived as helpful and, if not, how this could be improved.

As a first step we developed a questionnaire which was broadly discussed during the development. The survey included questions like:

- Did you use (a part of) QRiH when writing the self-evaluation? If yes, which parts? If not, why not?
- To what extent was QRiH helpful when writing the self-evaluation?

Furthermore, we asked the respondents to explain their answer.

We also included questions about specific parts of QRiH. For example: we asked whether the set of indicators for measuring research quality and relevance to society was helpful when writing a self-evaluation report. For every part of QRiH we asked the respondents to choose between: (i) this part was helpful, (ii) this part was not helpful and (iii) I don't know. We also asked them to explain their answer and suggest proposals for improvement.

The questionnaire for the survey was ready for use in the summer of 2018 and was distributed from this moment on until January 2019: the month when the last humanities research units in the Netherlands finished writing their self-evaluation. We sent the questionnaire to humanities research units shortly after they were finished writing their self-evaluation so that their experiences were still fresh in mind. The questionnaire was sent to 27 directors and policy advisors of the humanities research units, because they are usually closely involved in the process of writing and coordinating the self-evaluation. Some of them forwarded the questionnaire to people more closely involved with the self-evaluation. We received back a total of 21 questionnaires. There was a great variation in the way the questionnaires were completed. Some were filled out poorly, others extensively.

In order to get a better understanding of the experiences with QRiH, seven in-depth interviews were additionally held. The interviews were mainly held at faculties where the questionnaires did not provide a (good) coverage of the experiences with QRiH. The interviews were held with directors and policy advisors of humanities research units.

Finally, 16 self-evaluation reports of humanities research units were collected to study whether and how QRiH was used. In twelve of the sixteen self-evaluation reports QRiH had been used.

Preliminary results of the evaluation

In this paragraph we will first discuss the general results of the evaluation. Following we will address the results of three specific aspects of QRiH, namely the narrative, indicators and lists. A more detailed report of the evaluation will be presented later on.

General results

In general, we experienced a very high willingness to participate in the evaluation and also that most people were familiar with QRiH.

By far most respondents of the questionnaire think QRiH is very (5) or somewhat (12) helpful as an evaluation instrument. Many indicate that they are happy that QRiH has been developed specifically for the humanities. Besides this large group who believe QRiH is very or somewhat helpful, we also received a few neutral (2) as well as negative responses (1). The people who responded neutral mainly indicate that SEP works for them just fine and that therefore there is no need to use QRiH. We received one questionnaire of someone who is quite negative about QRiH. When sending out the questionnaire we received some negative responses, all from the same domain, philosophy. The main argument for them is that they already have a body to discuss quality indicators for their domain, and therefore don't need QRiH.

The narrative

The core of QRiH is that the entire self-evaluation is set up as a narrative. Almost all respondents indicate that they have used the narrative in their self-evaluation(s). Furthermore, the narrative appears to be the one component that people are most satisfied with. A number of the respondents indicate that the narrative (sometimes in combination with indicators) has helped them in finding the identity of the research unit (who are we?, who do we want to be?, how do we differ from other research units?).

The idea of using a narrative for writing a self-evaluation, in which the scientific and societal goals are described in mutual coherence, appeared to work quite well and is supported broadly. The evaluation also showed a few things of QRiH that could be improved. From the responses to the questionnaire and during the interviews, we discovered that not everyone had understood the idea of describing the *entire* self-evaluation as a narrative. For instance, some used a narrative only to describe the societal relevance of the research unit, basically following the SEP instructions. Besides this clarification, respondents indicate that the format for the self-evaluation should be adjusted on a few points, to make it more clear and that there is a need to share experiences and to see examples of narratives. All these comments and suggestions are quite practical and it does not alter the fact that most respondents and interviewees are positive about the narrative.

The indicators

The intention of QRiH is that research organizations choose an appropriate set of quantitative and qualitative indicators that provide robust data to support the narrative. QRiH has a wide range of indicators for production, use and recognition and are described in detail on the QRiH website. These indicators are the result of a broad consultation in the humanities field.

In the questionnaire, most people indicated that they found the set of indicators to be helpful. However, some respondents have specific comments on the indicators. Some indicate that there are too many indicators, others make suggestions for adding additional indicators. Some use the indicators of QRiH, others see the indicators primarily as an incentive to develop indicators for their own domain. Various respondents who indicated that they found the indicators to be helpful, explained that they were happy with the different types of indicators that represented the interaction of researchers with stakeholders (for production, use and recognition) and with the freedom to choose indicators that are suitable for them. The bottom up procedure in this respect was highly appreciated. Some also specifically express their appreciation for the indicator "hybrid publications", which stimulated the willingness of researchers to cooperate with the preparation of the self-evaluation. Furthermore, the self-

evaluations we studied show that, as QRiH intends, in many (but by no means all) self-evaluations is indicated why certain indicators have been chosen, i.e. why the chosen indicators are important for the research unit. It was also noticeable that many opted for quantitative social indicators, but a number (also) opted for more qualitative indicators, such as case studies.

The respondents and the interviewees regularly noted the practical problems with collecting data for societal indicators. For the collection of data for scientific indicators such as the number of journal articles or the number of citations, there are several well-known tools. For the collection of data for societal indicators, people are searching. There is a great demand for practical tips, instructions or simple tools for collecting, managing and analyzing data to support indicators of social impact. Furthermore, there seems to be a need to share experiences and tips and tricks. Some interviewees also mention that it is difficult to keep the data up to date over the six year's evaluation period also in view of the limitations of current research information systems. Systems such as METIS or PURE offer few opportunities to enlist the diversity of output allowed for in the QRiH system or to acknowledge the complex or hybrid characteristics of outcomes or the efforts that go with diverse output. Perhaps also as a result, staff may be consumed by other tasks, and collecting data loses priority.

Lists of journals and publishers

To help develop robust indicators, QRiH offers lists of prominent publication channels (publishers and journals). However, while a result of a broad consultation, these lists appeared to be the part of QRiH that the respondents are least satisfied with. The most important comments are:

- Faculties and other research units often cover several domains, while the lists are classified per domain. This is confusing.
- Not everyone agrees on the content of the lists. Some indicate that too little account has been taken of the subdisciplines and / or that not all relevant journals are part of the list. An interviewee indicates that it is impossible to make a good list, because this discipline is so wide and regionally different.
- Scientific domains have their own habits. Interviewees of one domain indicate that the lists are not used, because it is customary for them to look at journals with a high impact factor. For another domain, it is customary to use another list. Furthermore, there are people who indicate that the distinction between peer-reviewed and non-peer-reviewed journals is more important than the lists.
- One of the interviewees indicates that the problem with the lists is that they are seen as limitative, while this is not intended to be so. It can be reasoned (from the mission of the research unit) why publications are published in journals that are not on the lists.
- There was a lack of clarity in the process of the creation of the lists, which diminishes trust in the lists.
- Finally, there were practical problems: it turned out to be complicated for many to compare their own publication list with the QRiH list.

Concluding remarks

While we have not yet done a full analysis, these preliminary results will help us to develop QRiH further. The most important result for us is that there is a general appreciation for this initiative to find a way to evaluate humanities research that is closely related to the work people are doing, and to the kind of interaction they have with researchers in other fields, and

with the wider societal environment. The bottom up approach for indicators, qualitative and quantitative, to support the narrative in a robust way, was highly valued, not withstanding the fact that for many indicators still work has to be done. Especially the gathering of reliable data for indicators for societal impact is still difficult.

A controversial issue remains in the lists of journals and publishers. While these were meant to select the best or the most used in the field, it turned out to be a battlefield for different interests. The question is whether it is possible to compile a list that is recognized as a qualitatively robust representation of an entire field. Given the many comments on the lists, this requires further analysis and discussion with the field. A meeting for this will be organized in the autumn of 2019.

After that meeting, we will use the results of the survey, interviews and meeting to adapt QRiH for the coming years.

References.

Algra, K., Bennis, H., van Henten, J. W., & Kemperink, K. (2013). *Quality indicators for research in the humanities*. Royal Netherlands Academy of Arts and Sciences

Bensing, J., Andeweg, R., Franses, P., Meyer, B., Prins, C., & Schuyt, K. (2013). *Towards a Framework for the Quality Assessment of Social Science Research*. Royal Netherlands Academy of Arts and Sciences.

Prins, A., Spaapen, J., van Leeuwen, T., & van den Broek-Honingh, N. (2019). QRiH: Towards a fitting system for humanities research evaluation, *ftEval Journal for Research and Technology Policy Evaluation*, 48, 2019, p. 91-97

Prins, A. A. M., & Spaapen, J. (2017). Serving variegated audiences: From ranking oriented evaluation to mission oriented evaluation. *ftEval Journal for Research and Technology Policy Evaluation*, 44, 42-49.

Van der Hoeven, F. D., & Dijkgraaf, R. (2010). *Quality assessment in the design and engineering disciplines, a systematic framework*. Royal Netherlands Academy of Arts and Sciences.

van Vree, F., Prins, A., Spaapen, J., van Leeuwen, T., Duindam, D., & Boekhout, M. (2017). Handleiding Evaluatie van geesteswetenschappelijk onderzoek volgens het SEP.

VSNU, KNAW, NWO. (2014). *Standard Evaluation Protocol 2015-2021: protocol for research assessment in The Netherlands*. [https://www.vsnu.nl/en_GB/sep-eng.html]