

Manual

Evaluation of humanities research according to the SEP

21 September 2017



Manual

Evaluation of humanities research according to the SEP

Contents

- 0.** Introduction
- 1.** Evaluation according to SEP: basic principles and connections
- 2.** Aggregation levels
- 3.** Narratives as basic element of the self-assessment report
- 4.** The toolbox: QRiH
- 5.** Quality and relevance aspects
- 6.** SWOT analysis
- 7.** Case studies
- 8.** Conclusion: basic conditions and recommendations

Introduction

As regards robustness and methodological rigour, the humanities are in many respects no different from the sciences. At the same time, they have their own, recognisable academic culture and practices, of which the focus on a broader public rather than other academics is one of the key characteristics. The consequence of this is that researchers in the humanities can only make limited use of the systems and instruments that are employed a great deal in other academic domains to express the quality and relevance of research. In significant sections of the humanities - but also elsewhere - Web of Science and Scopus have no role to play, and this is also true of measuring instruments such as the H-index or citation scores. Neither do many humanities scholars feel that the one-sided focus on publications in English-language journals has much relevance to their work. In their domain, books, exhibitions and documentaries, including work in languages other than English, aimed at professional colleagues and a wider public, may actually be a better benchmark for quality, as it is in these areas that ground-breaking work often has an impact. Many researchers even see this as their task, because if they limited their efforts as researchers to the world of English-language peer-reviewed journals, they would cut themselves off from the living culture that is their source of nourishment.

However, all this does not mean that the quality of research results, in terms of academic and social relevance or impact, cannot also be evaluated in the humanities. The condition for this is that the assessment framework and the corresponding instruments used must lead to a judgement that humanities scholars themselves consider to be meaningful and valuable. In short, the evaluation must do justice to the values of the researchers in the various domains of the humanities¹, and enable all those involved to steer a course towards quality and relevance, beginning with the research groups themselves. We intend to achieve this by setting up panels in every humanities domain, which can determine the characteristics of quality and relevance. This assessment and reassessment of quality and relevance therefore takes place from the bottom up.

¹ In this manual two terms are used to refer to different areas of research: 'domains' and 'subdomains'. The term 'domain' denotes the humanities as a whole as well as the fields or disciplines that can be distinguished within that domain; specialisms and more specific fields are referred to as 'subdomains'. These two more or less neutral terms are preferable to 'disciplines' or 'fields', because in practice this terminology creates a lot of confusion and, what is more, it suggests clearly demarcated areas of academic activity - a suggestion that is in practice at odds with the multidisciplinary nature of much research. Only in the lists of journals is the term 'multidisciplinary journals' used, because it does not carry these connotations.

The most recent version of the *Standard Evaluation Protocol (SEP)*, which was drawn up by VSNU (Association of Universities), KNAW (Royal Netherlands Academy of Arts and Science) and NWO (Netherlands Organisation for Scientific Research) as a framework for evaluating academic research in the 2015 - 2021 period, offers ample room for making the evaluation toolbox suitable for diverse domains, not only in the humanities, but also in other domains where the above-mentioned one-sided, quantitative focus on publications in leading journals is under discussion. The SEP therefore gives the humanities the chance to position and present themselves comprehensively, not only in relation to academic quality, but also regarding their significance for society.

This manual has been prepared to enable research units at different aggregation levels in the humanities to write self-assessment reports that meet the SEP and at the same time do justice to the domain's own quality standards, as regards the valuation of research results both in academia and in society.

1. Evaluation based on the SEP: basic principles and connections

The SEP 2015-2021 wishes to enable a balanced assessment of both the academic quality and the relevance to society of work performed. The assessment is based on the self-assessment report, which is discussed on page 13 of the SEP and also in Appendix D, Format of self-assessment report. SEP Table D1 (output indicators) on page 25 forms the basis for providing evidence in support of the self-assessment report. The indicators in this table determine the form and content of the self-assessment report and the information gathered for the report forms the basis for the evaluation by the assessment committee.

| | | 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 |

Table 1 Types of output indicators according to SEP. Abbreviated version of Table D1: format of self-assessment report, p. 25

At first sight, such a self-assessment based on output indicators appears simple, but this is not the case. Firstly, it is tempting (and understandable) to fill this table with as much data as possible. However, this is not guaranteed to produce a clear picture of achievements, particularly as regards the connection between the academic and societal aims of the research.

Secondly, the indicators given are as yet empty categories, which can of course be used in a field-dependent fashion, but which in themselves do not say anything about the quality and relevance of the results that are collected together in those categories. In other words: the SEP does expressly distance itself from the overwhelmingly quantitative method of measuring quality, which focuses strongly on the exact sciences and is relatively one-sided. However, this does not alter the fact that the quality and relevance criteria that do justice to the uniqueness of the humanities and of the individual domains do have to be operationalised.

This manual addresses precisely this issue by:

1. providing a guideline for ordering and positioning the research results, with the emphasis on 'narratives' in which the various aspects of quality and relevance can be dealt with;
2. providing a set of indicators of quality and relevance that can be used in the various research domains of the humanities for writing a self-assessment report. These indicators can be both quantitative and qualitative.

The core of every self-assessment - regardless of whether it relates to a research group, a programme, a research institute or a faculty - is formed by what is referred to in this manual, in line with the terminology in the English-speaking world, as a 'narrative': an overarching, well-argued and substantiated description of the research unit and its mission and field of activity.

This approach therefore goes a step further than is formally prescribed by the SEP. In the SEP, narratives are stated only in relation to the evaluation of societal relevance. By applying the idea of narratives in a broader sense, greater emphasis is placed on the important fact that the humanities are about the connection between scientific and societal objectives.

Narratives as referred to in this manual show the broad objectives of the research unit, how it wishes to position itself in terms of forms of

production, audience, language and valorisation, and to what extent the unit has been successful in terms of quality and relevance. These narratives should be underpinned by robust data, organised according to the SEP indicator categories and, if desired, by case studies. This material as a whole provides the basis for the Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis, which, according to the SEP, is part of the self-assessment report and can function as a starting point for a policy geared towards quality and relevance. The decision to give self-assessments in the humanities primarily a narrative form articulates more precisely the flexibility offered by the SEP.

Narratives are substantiated, explicative stories. They need to be underpinned by robust data which serves as evidence - data that relates to the various categories of indicators listed in the SEP protocol. In this way narratives touch on each of the six dimensions of quality and relevance of research in the tabular overview above: the quality of the research (column 1) and its relevance to society (column 2), each considered from three points of view: the factual results or *output*, the use of this output and the recognition thereof (rows 1, 2 and 3). Narratives can also be underpinned by case studies, which in turn must of course be supported by robust data.

The indicators for determining quality and relevance, as shown in the six cells of SEP table D1 - with lists of examples, which, as stated by the SEP, can be expanded - are in part field-dependent. This manual explains how diverse indicators can be employed in assessing research in the domain of the humanities. On the Quality and Relevance in the Humanities website (www.qrih.nl), the indicators are worked out in detail (see sections 4 and 5 of this manual).

Self-assessment reports in the humanities thus consist of the following parts:

1. Introduction with short description, Profile and Ambition
2. Choice of relevant indicators
3. the results achieved in the domains of research and society
4. Conclusions of the self-assessment

These first four parts have the character of a coherent narrative, in which the character, profile and ambition of the unit to be assessed are described. This is then used as a background when dealing with the scientific and societal quality and relevance of the research results. Consequently, these first four sections constitute the core of the self-assessment report.

The structure of the rest of the self-assessment report follows the SEP format, although it should be noted that in any event that parts 6 and 8 are mainly given the form of a narrative.

5. (Administrative) context
6. Results of previous evaluations, SWOT analyses and future plans
7. PhD Programmes
8. Diversity
9. Research integrity, ethics, research data management
10. Robust data, a.o.
 - Case studies
 - Tables and appendices

Parts 1 to 9 of the self-assessment report consist in total of 15 pages; this is therefore excluding case studies, tables and other robust data.²

At the QRiH website this structure has been converted into an operational format for self-assessments, including an indication for the size of the various parts and an output table. The QRiH website also contains examples of self-assessment reports and of case studies.

This structure of self-evaluations in the humanities, with scientific *and* societal aspects being combined, offers a specific framework for the SEP. The reason for doing so is to provide greater cohesion and to reveal the dynamic within the research programmes.

2. Aggregation levels

Self-assessments, drawn up in line with the methodology outlined in this manual, can in principle be carried out at various aggregation levels - faculties, institutes, programmes and groups. Nevertheless, it is important to

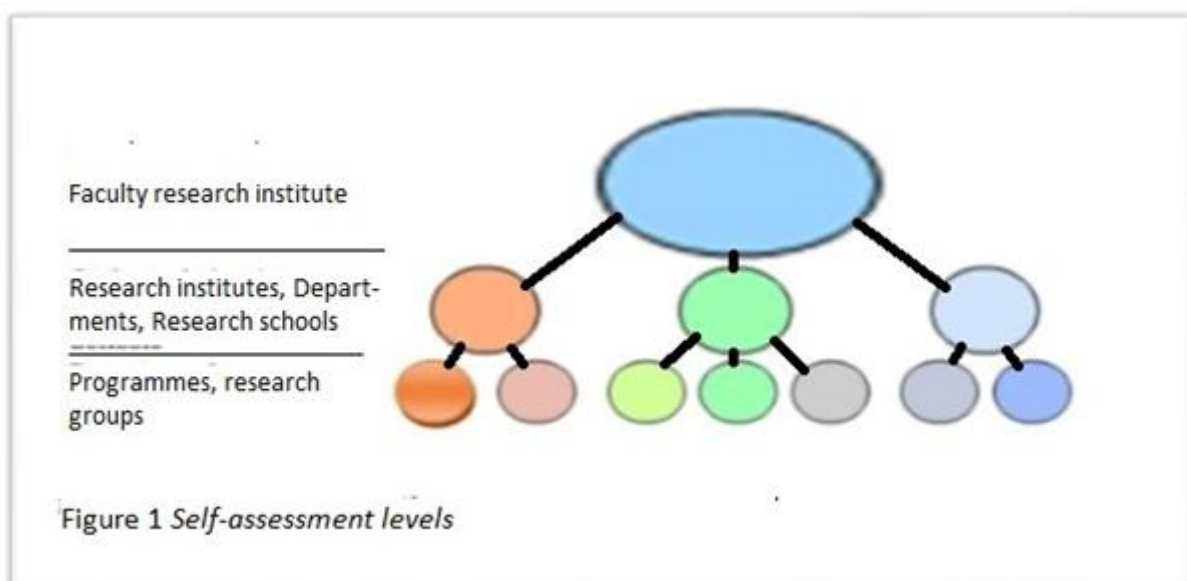
² For a comprehensive description of the structure of self-assessment reports, see the SEP, pages 23 and 24.

be aware that the character of the narratives and the use of indicators depend to a great extent on the level at which assessments are conducted.

The SEP manual (page 10) states that it is up to the responsible board to decide at which aggregation level the assessment will be conducted: 'the assessment will concern a research group, a research institute, a research cluster or the research carried out within a faculty'. The requirement here is that the units must have sufficient coherence and volume.

It is clear that there must be coherence: making assessments at the level of a research group or institute that is homogeneous as regards the discipline involved makes it possible to compile a clear-cut and unambiguous narrative, with robust data based on field-dependent indicators. The question is whether within the humanities, with its relatively large variety of disciplines, these units will always have the volume of at least 10 FTE of permanent research staff referred to by the SEP; in order to facilitate a high-quality and tailor-made self-assessment, there will need to be flexibility with respect to this prerequisite, with a view to the necessary homogeneity.

At the same time, it is quite possible to use the methodology described here for assessments at a higher aggregation level, in which results of research from several fields with different publication cultures are evaluated. This requires a more differentiated narrative than in the case of a homogeneous research unit, because only in this way can justice be done to the diversity of the different fields.



The requirement for differentiation is even more crucial in the case of self-assessments of very large and varied units, such as faculty-wide research institutes. To avoid such self-assessments being limited to the use of a set made up of a few general, shared indicators for quality and relevance – ignoring the principle of diversity – the recommendation is to draw up self-assessment reports at such a high aggregation level with the help of an 'overarching narrative'. Such an overarching narrative will inevitably be of a more synthesising and abstract nature, with the self-assessments of the underlying, more homogeneous units acting as the underlying robust data.

3. Narratives as basic elements of the self-assessment report

The SEP offers an overarching model in which – depending on the choices made by researchers and institutions, but also on the social context – a variety of legitimate research profiles are possible. These profiles are presented in a narrative form, which creates space for considerable variation.

As explained in the first section, narratives play a crucial role in the shaping of the various parts of the self-assessment report. In the narrative, the institute, the group or the programme indicates 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 narrative parts of the self-assessment offer plenty of opportunity to go beyond summaries of the results of projects and programmes. Such quantitative overviews are dealt with later on in the self-assessment report (part 9). The narrative parts focus on the research direction and ambition and on the related 'processes' that have been gone through. For example, the ambition of contributing to specialist knowledge through articles in international journals requires different efforts and also a different timeframe from activities such as developing and maintaining databases, creating exhibitions and catalogues, or writing monographs that are aimed at colleagues as well as being accessible to non-academic stakeholders (hybrid publications). The narrative can also look into the impact of past research over the longer term.

Narratives therefore offer scope for a description of the coherence of ambition and the diverse efforts that are made in the course of the research, including the effort to achieve academic quality and relevance to society, and the

process by which this coherence is created. In this way, narratives constitute the basic elements of self-assessment reports in the humanities.

4. The toolbox: QRiH

QRiH is the name of the toolbox on the Quality and Relevance in the Humanities website (www.qrih.nl). The website provides the tools for drawing up self-assessment reports in the humanities, according to the methodology described in the previous sections. The website contains profiles of the individual publication cultures, definitions of indicators and their possible applications, examples of applications, lists of indicators authorised by panels, examples of narratives and case studies, in addition to background information on the SEP and the QRiH toolbox developed.

The question that will face every research unit is which category of indicators it considers to be the most suitable to represent its work, and how to define the aspects of quality and relevance of its work. Therefore, the lists of categories on the QRiH website are not meant to be 'checklists', any more than the examples that are given in the SEP itself. On the contrary, it is not the list of indicators in the cells of the SEP that is meant to give direction to the self-assessment, but the narrative.

Which categories of indicators are especially relevant to a specific (sub)domain is determined to a great extent by the prevailing research and publication culture of the relevant research area or specialism. In this context, there are not only major differences in relation to the publication cultures in, for example, the scientific and medical domain, but also between research areas and domains within the humanities themselves. These differences specifically concern the channels of communication, with respect to the nature, form and language of the publication and of the target readership and the actual use of these products.

This can be illustrated with a few examples:

- in a large section of the humanities, monographs and edited volumes are the most important publication channels, also as regards the impact on the discipline itself;
- in a large section of the humanities, 'hybrid output', e.g. scientific publications, as well as exhibitions or reports, which focus simultaneously on researchers, clients and a wider public, is considered a very important communication channel, particularly due to the intended connection between science and 'living culture';

- publications in languages other than English can be of great value in a variety of disciplines, both as regards quality and impact.

Through its descriptions of different categories of indicators, together with their applicability and technical aspects, the QRiH toolbox provides the compilers of self-assessment reports with practical aids for structuring the report in a way that not only complies with this manual, but also does justice to the internalised values of quality and relevance in the individual disciplines.

5. Quality and relevance aspects

The SEP offers academic disciplines the opportunity to develop tools, or to have them developed, and to promote the use of robust indicators for quality and relevance. The aspects of quality and relevance developed below (with the corresponding elaborations on the website) together form a coherent toolbox for the purpose of compiling self-assessment reports in the context of external research assessments in the humanities according to the SEP. The basic principle is that in a self-assessment report all aspects of research that can demonstrate quality and relevance need to be described in a coherent fashion. For this reason, the tools for assessing quality concern not only direct results or outcomes, but also the processes that necessarily precede them or that are in another way intrinsically bound up with the outcomes.

Indicators for quality and relevance

The entirety of the data relating to the various quality aspects serves to underpin the narratives. The corresponding indicators constitute, as it were, a toolbox from which tools can be selected for the self-assessment reports. These indicators are systematically detailed and defined on the QRiH website, with an explanation of the possible use, the scope and, where necessary, the technical aspects. In view of the dynamics of the publication and research culture, the validity and usability of the indicators require continual testing and adjustment if needed.

The starting point is that the indicators used are the ones that are most appropriate for the work performed by the research unit. This does not alter the fact that a number of indicators are so characteristic of academic research (for example publications in academic journals) that they apply to all disciplines. In addition, there are indicators specific to particular fields.

It is up to the research units or institutions themselves to decide which indicators they want to use for their self-assessment. The choice of specific indicators should follow on logically from the prevailing research culture in the individual disciplines. Descriptions of the various research cultures can also be found on the QRiH website.

Authorised and reasoned indicators

Various categories of indicators of quality and relevance for the humanities are more ambiguous and also less developed than those for certain other disciplines, with the result that they are often judged according to criteria that apply to them either very little or not at all. Because of this, two forms of testing are employed to determine the robustness of indicators, resulting in 'authorised' or 'reasoned' indicators. Both categories can appear in each of the six cells in Table D1.

Authorised indicators are instruments that are established by representative panels of national research schools in the humanities and subsequently validated by the National Authorisation Panel (NAP), an authoritative and broadly constituted national panel, as is the case with lists of leading journals and publishers in a specific discipline. Where possible, the panels will act in accordance with international standards, such as the ERIH Plus lists or the Flemish (VABB) or Norwegian lists (CRiStin).³ As already stated, it is essential to have regular testing and adjusting of the authorised indicators.

Reasoned indicators are indicators that require sound argumentation regarding their use in testing quality. These categories of indicators include, for example, special publications, publications in languages other than English or Dutch, exhibitions, cooperation with civil society partners, etc. These reasoned indicators can also be of a quantitative nature, as will be shown by their description on the website. The research unit will itself have to draw up the arguments for the use of these indicators, and in a number of cases it can be possible to use standards developed elsewhere, such as the Flemish and Norwegian lists of qualified journals.

Indicators and interdisciplinarity

The assessment of interdisciplinary research demands particular attention. In a review article on this issue it is noted in summary:

³ <https://www.ecoom.be/en/vabb>; <http://www.cristin.no/>
<https://dbh.nsd.uib.no/publiseringsskanaler/erihplus/> Most recently consulted on 15 July 2017. The CRiStin is also used in other Scandinavian countries.

Existing output measures alone cannot adequately capture this process. Among the quantitative measures considered, bibliometrics (co-authorships, co-inventors, collaborations, references, citations and co-citations) are the most developed, but leave considerable gaps in understanding of the social dynamics that lead to knowledge integration.⁴

Most research domains in the humanities display a large degree of interdisciplinary interwovenness, although there is sometimes specialisation, for example in time periods, languages or geographic areas. This is true of cultural history, philosophy, archaeology or literature studies, as well as multidisciplinary domains such as media studies, medieval studies, cultural studies, religious studies and theology, and gender studies. In many domains there is also a substantive, methodological and theoretical kinship with social sciences such as anthropology and sociology. The connections with such domains are actively maintained, with the result that the research culture in one domain shows, in addition to specific characteristics, strong similarities with that of other domains. In the report *Kwaliteit en Relevantie in de Geesteswetenschappen* (Quality and Relevance in the Humanities) (2012, page 46), the KNAW warned that special care is needed in using internal (faculty) rankings for interdisciplinary research:

For this reason, when drawing up classifications one should choose an aggregation level that transcends the faculties. The suggestion is that lists that are already in circulation in various places should be reviewed in terms of what they have in common and what effects these lists have.

It is therefore also important that the panels be put together by the research schools, because these often have an interdisciplinary focus. They are also aware that peer review can have negative consequences for interdisciplinary research and they will therefore ensure that the panels are sufficiently broad in their composition.

6. SWOT analysis

⁴ Caroline S. Wagner, J. David Roessner, Kamau Bobba, Julie Thompson Klein, Kevin W. Boyack, Joann Keyton, Ismael Rafols, Katy Börner, 'Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature', *Journal of Informetrics* Vol. 5/1 (January 2011) pp 14–26.

The fourth - compulsory - part of the self-assessment report is the Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis. This analysis focuses on strategic prospects, and thus concentrates primarily on policy aspects and management aspects by describing the strong and weak areas of the unit to be assessed, in relation to internal as well as external factors. The SEP indicates in general terms how one can do this (page 30).

This part of the self-assessment report is in the form of a table containing the four elements of a SWOT, with a short explanation that explicitly looks at the future and considers the policy and the resources that are needed to strengthen the quality of the research. It goes without saying that the SWOT analysis builds on previous sections of the self-assessment report.

7. Case studies

Case studies, as referred to in the SEP, are an optional part of the self-assessment report. At the same time, they may play a crucial role, since case studies are excellent instruments⁵ to function as both illustrations and robust supporting elements of the first four sections of the self-assessment.

Case studies have a narrative form and may relate to particular projects or programs of the research unit, but also to certain aspects of the research activities, such as the interaction between research activities and society, or research and the PhD programs. Case studies may thus illustrate or highlight specific parts or aspects of the research, especially where it is considered important for the picture assessors may form of the unit to be assessed.

Case studies can be carried out at various aggregation levels: project, programme or unit as a whole. Whereas the SEP restricts the use of case studies to the impact on society, it is assumed here that case studies contain information about both the academic and the societal aims and output, and that they are pre-eminently suited to indicating the connection between the two - a connection that is seen as essential, especially in the humanities.

⁵ There are other ways of highlighting or incorporating illustrative elements into the narrative, such as working with text boxes: short case studies that describe the impact of the research or of a specific work or project on the academic domain or society.

In the 2014 British Research Excellence Framework (REF)⁶, 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. Examples can be found at (<http://impact.ref.ac.uk/CaseStudies/>).

Impact can take many forms. In the REF this is described as follows:

Impact is an effect on, change or benefit to:

- the activity, attitude, awareness, behaviour, capacity, opportunity, performance, policy, practice, process or understanding
- of an audience, beneficiary, community, constituency, organisation or individuals
- in any geographic location whether locally, regionally, nationally or internationally
- the reduction or prevention of harm, risk, cost or other negative effects.
- impacts within the higher education sector, including on teaching or students, are included where they extend significantly beyond the submitting institute.

Examples are given on the QRiH website.

8. Conclusion: basic conditions, guarantees and recommendations

This manual and the QRiH toolbox that has been developed enable researchers in the humanities to describe the quality and relevance of their research and have it assessed in a way that does justice to the values and standards in the academic domain in which they are active. To get the best out of this manual and the toolbox, research units must ensure a proper registration of research results. It will not be possible to produce robust data without this registration.

It should be noted that this instrument can be applied more widely than simply for assessments in accordance with the SEP. Experience shows that this manner of reporting is also eminently suitable in other contexts, for example in smaller research projects and programmes such as those funded by NWO or public-private organisations, among others.

⁶ The Research Excellence Framework (REF) 'is the new system for assessing the quality of research in UK higher education institutions. It replaced [in 2014] the Research Assessment Exercise (RAE), last conducted in 2008 [...] The assessment provides accountability for public investment in research and produces evidence of the benefits of this investment'. See <http://www.ref.ac.uk/>

More important still is that this tool can be used to shape research policy at various aggregation levels, whether or not this is tied in with the SWOT analyses. This could take in human resources policy, for example to shape diverse objectives in terms of quality and relevance at the level of research groups or teams.

Finally, the toolbox that has been developed, in all its versatility as regards quality and relevance, could be of help to individual researchers in obtaining a picture of their work environment and thus of their possibilities and career prospects.

To be able to use this instrument for the assessment of quality and relevance in the humanities for future occasions, it is vital that the national organisations in the area of the humanities, notably the national boards of faculties of the Humanities (DLG), Philosophy (DWB)) and Religious Studies (DGO), and the Royal Academy of Arts and Sciences (KNAW), who set up this tool as the *Partnership QRiH* (Samenwerkingsverband QRiH), as well as the national research schools, commit themselves to this system and act as guarantors for its maintenance and further development.

The *Partnership QRiH* has installed a small-scale organisation to support the National Authorisation Panel, the tasks and powers of which have been laid down in regulations (see the www.QRiH.nl website). It is the task of the NAP to ensure the updating and possible revision of the authorised indicators, such as the relevant academic journals and publishers, but also adjusting and further developing or making operational other indicators that have possibly not yet been authorised, all in cooperation with the domain panels and the working group of experts.
