

ALIGN WHERE POSSIBLE, UNIQUE AND SELF-CONFIDENT WHERE APPROPRIATE

DEVELOPMENTS IN LEGAL RESEARCH AND IN ITS ASSESSMENT

Overarching report of the 2016-2017 research assessment cycle
Prof. Willem van Genugten, June 2017

The report has originally been written in Dutch and is translated by Rosenbrand Vertalingen, Maarsse.

In 2016, the research at the Dutch Faculties of Law was reviewed.¹ In contrast to previous years, each Faculty was assessed individually by a separate review panel, rather than all legal research being assessed by one single national review panel. This choice was made by the Law School Deans in joint consultation, using the option offered in the VSNU Standard Evaluation Protocol. It was inspired by the strongly-felt need for more in-depth reflection regarding the concrete research results and individual research profiles and strategies, an effort that would be too comprehensive for a national review panel. The Council of Law School Deans asked Prof Dr Van Genugten to act as an intermediary between the various panels in order to promote equality in the evaluation and assessment. Prof Dr Van Genugten was also requested to offer his reflections on the final findings of the review panels and to present an overview of the state of affairs in the national and international world of academic legal research, of the most striking current developments, and of the discipline's position in relation to other relevant academic disciplines. Prof Dr Van Genugten was given the opportunity to do so on the basis of his own views as inspired by his knowledge and experience. This report presents his findings and considerations.

The Council of Deans was pleased to receive this report. The Dutch Law Schools frequently need to change orientation to respond to both academic and social developments. In doing so, the discipline takes into account the methods of legal research, increasingly considering the methods and insights of other disciplines in order to enhance and broaden its own insights. At the same time, legal research is well aware of its social role and responsibility. Its contribution to strengthening the legal order in modern society is essential and critical.

The present report by Prof Van Genugten offers the Faculties of Law some very practical ideas for the current debate on how legal research in the academic world and beyond should further develop and strengthen its position both nationally and internationally.

On behalf of the Council of Law School Deans

Prof Dr A.M. Hol
Chairman of the Council of Deans

1 At Utrecht University, the School of Law is part of the larger Faculty of Law, Economics and Governance.

Approach and acknowledgements

It will not have escaped the notice of anyone in the field that recent decades have seen numerous publications on what makes legal research a science and what are the best ways to measure such research and decide on the appropriate qualitative indicators. The underlying questions here are what can legal research and those responsible for assessing it learn from other disciplines, and what specific academic qualities and measurement tools should they use?

This report does not set out to reflect on the present status of this debate – assuming that there is actually a specific status. Instead, it aims to contribute to the debate by conveying observations from the field and from my role as national reviewer/coordinator of the 2016-2017 assessment of academic studies in the field of law and as an observer during the assessment visits (with the exception of Tilburg). It also builds on the final reports of the nine decentralised faculty assessment committees and on discussions within these committees, which together form the essence of the system chosen for the present assessment cycle. As part of this system – which is discussed in more detail in section 1.1 – the deans of the nine law schools opted for an approach not focusing, or at least focusing to a lesser extent than previously, on comparisons between the different schools and research groups, and that instead concentrates more on the extent to which the faculties have achieved the ambitions they set for themselves with regard to the core assessment criteria of ‘research quality’, ‘relevance to society’ and ‘viability’.

Similarly, this report has *not* been written for the purposes of ranking faculties. Instead, its primary aim is to provide insight into the present position of legal research and to position this research alongside and in the midst of other academic disciplines. I have opted in this respect for a helicopter view, with specific consideration for the assessment process and for faculties’ research and research policies, as assessed by the nine committees, while at the same time seeking to concentrate primarily on their similarities rather than on their differences.

When being approached as reviewer/coordinator, the deans specifically asked me to use my experience as a dean, research leader and researcher (see the annex for my CV). In addition, I asked the nine committee chairs (see section 1.2.1 for their names), together with two former law faculty deans who were closely involved in preparing the assessment (Professors Rick Lawson from Leiden University and Corien Prins from Tilburg), to read the draft report to ensure that the views I express are generally accepted as conveying a realistic picture, without of course asking them to agree with me on a detailed level. I am very grateful to them for their reactions.

Although this report does not contain any footnotes, all the points made can, if necessary, be traced back to a variety of sources. The report also regularly includes elements from the evaluation protocols. For reasons of readability, however, I have generally summarised these elements in my own words and omitted less relevant nuances. However, it should be emphasised that ‘no rights can be derived’ from this approach and that, in the event of protocol sensitivities and subtleties of meaning, the wording in the original texts will prevail.

CONTENTS

Central thread and recommendations	1
1. ASSESSMENT PROCESS	4
1.1 Framework	4
1.2 Actual performance of the assessment	5
1.2.1 Nine committees and their support	5
1.2.2 Consultations	6
1.2.3 What had to be assessed?	7
1.2.4 How were the assessments conducted in practice?	8
2. THE STATE OF THE ART OF LEGAL RESEARCH	11
2.1 Issue-driven and ‘wanting to matter’	11
2.2 Pluralism of methods	12
2.3 Variety of publication forms and fora	13
2.4 Relevance to society	13
2.5 Internationalisation	15
3. ORGANISING LEGAL RESEARCH	15
3.1 Programmes and group sizes	15
3.2 Position and role of PhD candidates	16
3.3 Instruments for steering and control	17
3.4 Interaction between contents and organisation	18
4. QUALITY AND MEASURING QUALITY: REFLECTIONS ON VARIOUS INDICATORS	19
4.1 External research grants	19
4.2 Third, fourth and fifth flows of research funding	20
4.3 External peer reviews and assessments by editors	20
4.4 Impact factors and citation indexes	21
4.5 Dealing with concurrence of academic research and practical relevance	21
4.6 Relevant positions and significant invitations	22
4.7 Membership of editorial boards	22
4.8 Knowledge for policymaking	22
4.9 In conclusion	22
5. BENCHMARKS	23
5.1 Actual use of benchmarks during current assessment	23
5.2 Benchmarking: looking ahead	24
6. CONCLUDING OBSERVATIONS	25
6.1 Three consecutive research assessments	25
6.2 Similarities and differences between the faculties	26
6.3 Diversity	27
6.4 Integrity	27
6.5 Awarding ratings	28
6.6 Benefits of the self-confidence of the discipline	29
ANNEX Author’s cv	30

Central thread and recommendations

Traditionally, assessment reports do not include summaries. As committees carefully deliberate and weigh up their words, reducing those carefully chosen words to an 'extract' or 'executive summary', let alone to the numerical ratings accompanying the assessments, does do a disservice to those deliberations. It also is too easy for the ratings to overshadow the substantive information in the assessments. This overarching report does not include a summary either, for two reasons of a dissimilar nature: a) We live in an age in which the excessive role of protocols and their statistical, quantifiable and quantitative outcomes are being discussed in numerous sectors of society, including academia in general and the field of legal research. These discussions frequently end with a sigh ('We've got no say in it. So nothing we can do about it' or 'That's just how it is these days') and then it is back to *business as usual*. A summary would detract from something of principal importance to this report: the need to fundamentally reflect on the position of legal research anno 2017, in close interaction with the management and organisational tools available to the various parties. b) The report is targeted at a varied public, and what is important to one research group may not necessarily be important to another. The same goes for boards of schools and universities, as well as for universities' external readers.

Rather than providing a summary, therefore, I have sought to provide the reader with a detailed table of contents, with meaningful subheadings, many cross-references within the text and a central thread running through the report, all with the aim of presenting a readable report, while also saving readers' time.

The central thread running through the report is as follows:

- The report is not arguing against protocolisation. In my view, periodic reviews of legal research based on well considered, albeit initially not always ideally suitable protocols have greatly benefited the discipline. Assessments have proved to be a tool that makes legal scholars think about 'what we are doing', 'how we are doing it' and 'who we are doing it for'. The various assessments over the years have represented a challenge to the discipline and, to a certain extent, forced it to think about such issues in a structured way and to act in a manner reflecting the outcomes of all these considerations. However, the report suggests in various places that protocols and quality indicators should remain primarily a *tool* and should in no way replace substantive discussions of quality.
- The nine chairpersons and the author of this report believe that the current 'coordinated decentralised model', which is very much focused on 'learning' and not on 'competing', at least not *between* the schools, has generally worked well across the board. And it is all the easier for them to say this, given that none of them was involved in deciding on the method to be used for the present assessment. As prescribed by the Standard Evaluation Protocol ('SEP') 2015-2021, the actual assessments focused on 'research quality', 'relevance to society', 'viability', 'PhD programmes c.a.' and 'research integrity', to which a great deal of purely substantive attention was devoted during the discussions. Last but not least, considerable attention was also devoted to the ambitions of the law schools and their research groups, their ability to learn and, by extrapolation, their visions for the future. The combination of the above resulted in nine separate reports and this overarching report. Together these can help anyone involved in legal research to reflect on the present state of the discipline and on the ways forward, both with regard to substance and to internal and external quality control.
- In contrast to previous assessments (Van Gerven, 1996; Ten Kate, 2002 and Koers, 2009), it was consciously decided by the deans of the nine law schools not to structure the assessment as a national comparative and rankings exercise. As stated earlier, the same essentially applies to this report. Like the nine faculty reports, this report also contains figures, although not

comparing faculty or research group level (see section 1.2.4). In addition, the report reflects on the awarding of ratings (see section 6.5). Readers wanting to know the ratings awarded to each law school and, if applicable, to each research programme can find this information on the internet as the schools and the executive boards of the universities have to publish their assessment reports and their responses (at least their initial responses) within six months after the site visits. Although those interested can then compare the information, they will see that the ratings provide only a very limited basis for comparison. From the perspective chosen for the present assessment, that represents a step in the right direction as it forces readers, more than ever before, to look at the underlying ‘narratives’ and the arguments and considerations expressed.

- In recent decades, the output of legal research has often been assessed on the basis of criteria from the ‘exact sciences’, without taking proper account of the specific characteristics of the legal discipline itself. At the end of the report I conclude that the tide has now turned in favour of a.o. the legal discipline and that the SEP has moved closer to the world of legal research rather than the other way round. Instead, however, of this constituting a return to the situation of some decades ago, I like to emphasise that this points more to an embracing of discipline specificity, and this – obviously – extends beyond legal research alone. Similarly, I do not see the (title of) the report as giving any cause to turn our back on other academic disciplines (‘Align where possible’), while at the same time there are also many reasons to stand up for and defend the specific characteristics of legal research (‘Be unique and self-confident where appropriate’). Meanwhile, the subordinate title of the report also refers to ‘developments’, both in legal research itself and in the way it is assessed. And this evolutionary process will always be ongoing.

Generic recommendations for future assessments

- The report as a whole should be taken as a recommendation, where necessary, to continue reflecting on the nature and substance of legal research, its relevance to society, the viability of faculties and research groups, the appropriate academic and organisational leadership and the various tools available to the discipline in this respect. I also make a number of practical suggestions, specifically in part 4, for further improving the way one can measure the quality of legal research. In most cases, however, these suggestions are preceded by raising a question about the friction, or potential friction, between ‘refining quality control’ and ‘more bureaucracy’. This is illustrated in the following example, highlighting the contrasts between the two extremes. During the assessments it has become clear that the individual faculties handled the terms ‘peer review’ (i.e. external peer review) and ‘refereed’ (i.e. the official SEP terminology) somewhat differently. As a result, this report contains a number of recommendations for clearer, unequivocal use of such crucial terms, specifically focusing on the overlap between external peer reviews and appraisals by strong editorial boards, possibly supported on an *ad hoc* basis by experts in sub-disciplines (see section 4.3). Looking ahead to future assessments, consideration could also be given – and that is not stated in this report itself – to appointing a small national committee to perform random checks with regard to the actual use of ‘peer review’ c.a., for instance a year before the next assessment cycle, assuming that the risk of having to undergo an external control of this nature would in practice already serve to sufficiently ‘sharpen minds’. It would not, however, be appropriate for this report to run ahead of itself in recommending such changes. This is up to (future) boards.
- This also applies to the form of the next assessment. The nine chairpersons and I see little reason to return to a system of national assessments by a single committee, even if that were to be feasible. We all know the two systems from inside out and, without being able or wanting to make an in-depth study of the pros and cons of each system (see section 1.2.4 for some

arguments), it is clear to us all that the current approach has produced quite some benefits and that the variation was in any event a good idea. We believe, however, that even if the current model is to be retained, one should seek to further standardise the approach, starting with timely agreement on and consistent application of critical input criteria such as the ones referred to above, given that these form the basis of any assessment. Looking to the future, one could conceivably also devise a 'hybrid' method, such as arranging for all private law research in the Netherlands to be compared by a single committee, and so on. In that case, the assessment would be a national assessment, structured around individual disciplines. That would be possible, and maybe other disciplines have already used such a system. However, this, too, would first and foremost require an executive decision on returning to a national comparative model, whether in part or whole. Whatever the case, all sorts of variants are conceivable.

- Having read the findings in the relevant faculty reports and this report, any self-respecting director, research leader and research group will feel called on to further reflect on the way forward, including any action to be taken. Assessments are ideal moments for this, given that day-to-day routines quickly tend afterwards to regain the upper hand. At the same time, it is important that such introspection should be done within a reasonable period (six months or a year), irrespective of the individual roles within this process, as the more promptly and effectively action is documented, the more progress – or lack of progress – will be visible and able to be monitored internally and externally later on. The fact that mid-term reviews, and all the reporting burdens that these involve, are no longer compulsory may help in this respect as it should avoid faculties and research groups feeling they are in a permanent state of external supervision and control. Let that be even more of a reason to encourage the discipline to think about reinforcing internal, content-driven controls of quality and focusing on both content and the organisation in the future, while also allowing time for specific priorities and deciding what *not* to spend time on ('posteriorities'). This would seem to be the best preparation for the next assessment, which will come, whatever form or hybrid form it may take.

1. ASSESSMENT PROCESS

1.1 Framework

Until the end of the last century the Association of Universities in the Netherlands ('VSNU') was responsible for conducting periodic assessments of university research and education. Since the turn of the century, however, this responsibility has been shared with the Netherlands Organisation for Scientific Research ('NWO') and the Royal Netherlands Association of Arts and Sciences ('KNAW'). Together, these three organisations drew up the Standard Evaluation Protocol ('SEP') 2003-2009, followed in 2014 by the SEP 2015-2021. In principle, the SEP forms the basis for assessing *all* academic disciplines, while also leaving scope for discipline-specific additions.

The SEP focuses on three core assessment criteria: research quality, relevance to society and viability. Given that the present assessment is based on the SEP 2015-2021 and is applied with retrospective effect, and also given that faculties will have tailored some of their activities and behaviour in line with the old SEP, it is important to note some changes in the new SEP. The most important of these are undoubtedly that productivity is no longer an *independent* criterion, while more attention is now also being devoted to societal relevance, and while the strategy and ambitions of research units that are being assessed are now regarded as leading elements in the assessment process. Other changes include the minimum size of the research units being assessed, the increased focus on PhD programmes and candidates, the policy on research integrity, and the introduction of a new rating model. The model previously applied was based on ratings on a scale of 1 – 5, with 5 as the highest rating and the opportunity to award ratings in intermediate categories, whereas the new model comprises ratings on a scale of 1 – 4, with 1 as the highest rating and no opportunity for ratings in intermediate categories (for more information, see section 6.5).

The present assessment was based on the generally applicable SEP standards, as transposed and detailed in the 2016 Discipline Protocol for the Assessment of Legal Research ('the Discipline Protocol') drawn up by a working group chaired by Prof. Fabian Amtenbrink, Director of Scientific Research at the Law School of the Erasmus University, Rotterdam. The Discipline Protocol was adopted by the Council of Law School Deans in 2015. The Protocol relates in particular to the indicators on which measurements of research quality, relevance and viability are to be based, with the first aim being to align with generally applicable indicators such as refereed articles, external grants, dissertations, visiting professorships and invitations to give key note speeches, as well as, secondly, to identify specific indicators such as annotations, manuals and advisory work in legal practice. More information on the indicators can be found in section 1.2.4 (for the present assessment) and part 4 (looking to the future).

With regard to the framework for the present assessment and within the opportunities available under the SEP, it is also crucially important to note that the deans jointly opted for decentralised assessment committees, primarily to reflect on specific characteristics of the nine participating faculties and without seeking to focus on individual comparisons and rankings. This contrasts with the approach adopted in 1996 ('Van Gerven Committee'), 2002 ('Ten Kate Committee') and 2009 ('Koers Committee'). At the same time, the faculties were allowed considerable freedom to determine the *scope* and *aggregation level* of the assessment for themselves. The choices made were as follows:

- Assessment of the faculty as a whole: Amsterdam (UvA), Maastricht, Rotterdam and Tilburg;
- Assessment of the faculty as a whole plus the individual research programmes: Amsterdam (VU);
- Assessment of the faculty as a whole plus the individual research programmes, with a qualitative appraisal of the faculty as a whole and an appraisal including ratings for the individual research programmes: Leiden, Utrecht and Nijmegen (NB: in the case of Nijmegen, only public law

research was assessed, given that the private law research had previously been assessed as part of the KNAW reaccreditation procedure);

- Assessment of the faculty as a whole plus the individual research programmes, with an assessment including ratings for the faculty as a whole and a qualitative assessment without ratings of the individual research programmes: Groningen.

The law faculties of the two Amsterdam universities and the universities of Maastricht, Nijmegen and Rotterdam also took the opportunity to submit additional questions on faculty-specific points to their committees. These are not discussed in any more detail here.

The research conducted by the existing national research schools has not been included in the assessments as this is covered by separate procedures, while the contributions of separate law schools to these schools has been inserted in the self-assessment reports. The Open University's research will be assessed in 2018.

In order to ensure some degree of uniformity in the assessment process, the deans also appointed the undersigned as a national reviewer/coordinator with responsibility for: a) Subjecting the composition of the decentralised committees to a marginal test of compliance with the relevant criteria in the SEP and the Discipline Protocol (see section 1.2.1); b) Attending the committees' site visits as an observer; c) Preparing an evaluation report on the present position of legal research in the Netherlands, including reflecting on the assessment methods currently applied. The combination of these tasks is referred to in the Discipline Protocol as 'coordinated decentralisation'.

1.2 Actual performance of the assessment

1.2.1 *Nine committees and their support*

The nine decentralised assessment committees and the combinations of members selected had to meet the following SEP requirements: inclusion of international members; familiarity with recent trends and developments in the relevant research fields (also in an international, comparative context); ability to assess the applicability and societal relevance of the research; a good knowledge of the Dutch research culture, including the external funding mechanisms and PhD programmes; a familiarity with the policy on research integrity.

The Discipline Protocol additionally requires "the ability within the committees to assess research focusing on Dutch law and/or Dutch-language publications" (p. 5). The SEP also stipulates that the committee secretaries should not be associated with the universities to which the faculties being assessed belong and should also have professional experience of assessing academic research, including the related procedures.

The nine assessment committees were chaired by Prof. Leonard Besselink (chair of the Maastricht committee), Em. Prof. René Foqué (Nijmegen), Prof. Laurence Gormley (Rotterdam), Dr Alfred Hammerstein (Leiden), Prof. Paul van der Heijden (Tilburg), Em. Prof. Teun Jaspers (Groningen), Prof. Pierre Larouche (UvA), Prof. Taru Spronken (Utrecht) and Prof. Paul Torremans (VU). As well as their academic stature, each of them has experience in legal practice, advisory work or managing universities and faculties, or a combination of these areas. The visits were made between 3 October and 13 December 2016.

In total, the committees comprised 50 members (including one person with a dual role), of whom 25 were from outside the Netherlands (including some professors of foreign origin, but holding a position at a Dutch law school). Further to this, the composition of the committees showed quite some diversity. To illustrate: ten members are currently professors at Dutch law schools, while five are emeritus professors and again five other members had a background primarily in legal practice in a broad sense. In retrospect, and as a somewhat late side remark, it would have been interesting

to have applied the diversity criterion, which was not added to the SEP until 2016 (see section 6.3), when appointing committee members.

With regard to the ten ‘current professors’, the decentralised nature of the assessment made it easier than before to approach potential committee members who are currently professors at Dutch law faculties. This was done quite intensively (one in five of the committee members), obviously in the hope and expectation that committee members would learn from each other and so, either intentionally or as a side effect, promote cross-fertilisation within the legal discipline as a whole. Five of the nine committees were chaired by professors currently working at one of the nine faculties, while two of the nine chairpersons (Hammerstein and Jaspers) were also members of the Koers Committee (2009).

1.2.2 Consultations

In late September 2016, and so before the first assessment visit, the nine chairpersons and the undersigned met for consultations in Utrecht. At this meeting, they discussed the contents of the SEP and the Discipline Protocol and how to deal with them in practice. The three most important conclusions and agreements were as follows:

- The SEP and the Discipline Protocol reserve the rating of 1 (equivalent to ‘world-leading/excellent’) for research units demonstrating the following three qualities: a) The research unit has been shown to be one of the few most influential research groups in the world in its particular field; b) the research unit makes an outstanding contribution to society; c) the research unit is excellently equipped for the future.
These standards have to be met cumulatively, with any rating of less than 1 for any of the three elements making it impossible to be awarded an overall rating of 1. The Discipline Protocol states that research programmes “primarily focusing on national law”, other than law research groups and programmes “genuinely active in research fields in which competition is international” (p. 8), are not eligible for a rating of 1. The Utrecht meeting decided, however, that achievements by groups with a national focus should in principle be able to be awarded such a rating. Key words in this respect are ‘exceptionally good’, ‘leading’, ‘innovative’, ‘relevant to society’ and ‘viable’; in other words, the equivalent of ‘excellent’, but then on a national scale.
- As agreed between and with the deans, the committees would focus on conducting open and frank dialogues and on the idea of the faculties holding up the proverbial mirror to themselves. In this, the *guiding principles* would be the three key assessment criteria in the SEP (research quality, societal relevance and viability), while the *object* of the assessment would be the self-assessment reports that the faculties themselves compiled and the information they added to them, either orally or in answer to questions posed by the committees, and the *instruments* would be a critical appraisal of the available material against the standards set. The standards and indicators used would be as laid down in the two protocols, but the appraisal would also and primarily be based on the wording used by the faculties and their research groups in their own self-assessments to demonstrate their ambitions (including the strengths and opportunities in their SWOT analyses), and the way they present themselves during the site visits. The meeting in Utrecht also discussed the opportunity to diverge during the visits from the agreed programmes so as to allow scope for additional discussions or, for example, to allow discussion partners to be divided into smaller groups to address issues that might otherwise be discussed only superficially. In practice, this happened on several occasions.
- It was also agreed to incorporate a moment of reflection into the assessment procedure after the nine visits had been completed. This discussion, which took place early February 2017, was also held in Utrecht. Before that, in mid-January, I (alone) received each of the nine draft reports and then raised some general points and observations, which could not be traced back to

specific faculties. The consultations resulted in an open and frank exchange of views on parties' experience of the process and ended with my requesting the chairpersons to take another look at the draft reports in the light of what we had discussed. Given that the nine committees were autonomous, my coordinating role ended at that point, and I did not seek to establish whether any changes were subsequently made.

From a coordination perspective, it should also be noted that the committee secretaries regularly exchanged e-mails on, for example, the key elements and structure of the reports (within the frameworks outlined in Appendix E to the SEP) and, in liaison with me, on the annexes to be included (i.e. showing the number of research FTEs, output, flows of funding, and PhD candidates). This all helped to promote and ensure 'a degree of uniformity in (what is still a great deal of) diversity'.

1.2.3 What had to be assessed?

The executive boards of the participating universities had to check whether the unit being assessed has at least 10 research FTEs (excluding PhD candidates and post-docs), has existed for at least three years, is recognised as a unit internally and externally, and whether it could include a suitable benchmark in its self-assessment. Research units with fewer than 10 research FTEs were also eligible for an assessment, providing the faculties had properly substantiated reasons for this and these were approved by the relevant executive boards. In practice, this happened on a considerable scale, with 26 of the total of 30 research programmes for which a general or qualitative assessment was requested having fewer than 10 research FTEs (excluding PhD candidates and post-docs) in the reference year of 2015, while 14 of these 26 programmes have fewer than 5 FTEs.

Section 1.1 refers to the *aggregation level* chosen by the faculties. It is also important in this respect to note that, irrespective of their individual choices, the faculties had to submit *all* their research to an assessment, including research conducted outside research programmes and research for educational aims. Although PhD candidates and post-docs were admittedly excluded in calculating the numbers of research FTEs, the contents of their work and how, for example, their programmes are structured did have to be included in the self-assessments and to be discussed during the assessments.

Under both protocols, the assessment of the *content quality* of research performed had to be based first and foremost on the key publications presented. In contrast to the SEP, which assumes the presentation of five scientific publications and five publications of societal relevance per unit, the Discipline Protocol opts for one key publication per research FTE (again, excluding PhD candidates and post-docs), with an explanation as to why each publication is regarded a key publication and whether it is primarily of scientific or societal importance, or a combination of the two. The latter touches on an aspect discussed in more detail later on in this report, being the actual or potential concurrence of scientific and societal significance in legal publications (see specifically sections 2.4. and 4.5).

With regard to *relevance to society* as a quality criterion, the SEP refers to contributions targeting specific economic, social or cultural groups and to advisory reports and contributions to public debates, adding that a research unit can specify contributions in areas that the unit itself has designated as target areas. Owing to the scope allowed to faculties and research groups – once again, it should be noted that the SEP covers *all* disciplines – the SEP gives just a few examples of indicators that can be used for assessing research's relevance to society. Examples given under the heading of societal relevance in the Discipline Protocol include authoring specialised publications and manuals for use in legal practice, participating in public debates (in the media or in blogs, for example) or being a member of an advisory body, and otherwise leaves considerable scope to the faculties and research groups to decide on indicators of their own. The Discipline Protocol also notes that law research is often intertwined with other academic, professional and popularising activities,

and that research units can therefore also opt to present a *showcase* in which all the various aspects of their work, as it were, come together organically.

With regard to assessing *viability*, the third key quality criterion, the SEP states that what matters is the strategy that the research unit intends to pursue going forward and the extent to which it is likely to be able to achieve its targets. The key words here are: the viability of the topic, academic leadership and management, staff and financing, together comprising the entire chain of factors and actors needed to achieve the ambitions the unit has set for itself.

With regard to the assessment of *PhD programmes*, for which the SEP and the Discipline Protocol request special attention, the crucial aspects include selection and admission procedures, programme content and structure, supervision, institutional quality assurance, duration of PhD studies, numbers of candidates failing to complete and career prospects.

In the case of *research integrity*, the SEP and the Discipline Protocol refer to matters such as the existence of a policy on the subject and researchers' sensitivity to integrity issues, the transparency of the research culture, interaction between management and researchers and among researchers themselves, and the extent of critical self-reflection within the institution or unit. Both protocols also state that matters such as data management, including the handling of raw data and processed research data, along with protocols for storing such data, should be taken into account.

Under the SEP, and partly in order to reduce the assessment workload, the faculties are no longer obliged to subject their research to mid-term reviews, whereas in the past, and as required in the SEP 2003-2009, all the faculties arranged for such a mid-term review. As part of the procedures for the present assessment, the faculties were asked to make the results of these mid-term reviews available to the assessment committees and to reflect on the outcomes in their self-assessments, albeit that the nine committees were completely free to reach conclusions on these reviews as they saw fit.

1.2.4 How were the assessments conducted in practice?

Crucially and wholly in line with the freedoms allowed in the SEP, the deans opted for a model in which strong, but collegially minded committees, chaired by individuals with a proven track record, entered into a debate with the faculties on their research performance in the years 2009-2015 (and often taking account of subsequent developments, albeit that these were sometimes contained in the narrative sections of the self-assessments rather than in the data lists, as inclusion in the latter was not permitted). These discussions were open and frank, generally starting from the ambitions that the faculties had set themselves and with the key concepts in the SEP and the Discipline Protocol (i.e. quality, relevance to society and viability) as the guiding principles for the discussions, along with the information supplied as 'test material' on the basis of the indicators set. This was the starting point for and also how the discussions proceeded.

The debates reflected the 'state of the indicators at that moment in time'. The idea underlying the SEP and the Discipline Protocol is that the assessment of the research results, relevance to society and viability should be objectivised to allow the performance achieved to be established as conclusively as possible, both internally and in comparison to comparable faculties/research groups elsewhere in the Netherlands or abroad. Two particularly important key words in this respect were 'benchmark' and 'indicators'. I will discuss the benchmarks separately at a later stage (in part 5). As to the *indicators*: as noted earlier, both protocols specify a whole range of indicators, while also simultaneously leaving scope for the use of discipline-specific or research unit-specific indicators. And although the Discipline Protocol is very clear in certain respects and streamlines matters, and while all the faculties submitted data in accordance with the Protocol – varying in weight, to put it in other terms, from some 400 grams to over 7 kilos – it became evident, once again, that it was not always possible to draw hard and fast, unequivocal conclusions on the basis of the material submitted.

Reference should be made in this respect to three factors relating, respectively, to definitions and the way they are used in practice, to incomplete files and to the way faculties present themselves:

- Sometimes it was not easy to establish how *exactly* faculties interpreted terms such as ‘refereed publications’ and ‘scientific publications’. In general, this was obviously clear, but more specifically the definitions used were not always entirely precise in terms of providing a basis for qualitative and quantitative conclusions (comparative or otherwise). The same applies, for example, to the attributing of externally obtained grants to the second or third flow of funds. Although it could be queried whether that makes much difference in day-to-day practice, it could potentially be a problem from an assessment perspective, given that the second flow of funds is associated primarily with external recognition and the third flow of funds with relevance to society, or a combination of the two.
- Not all faculties systematically recorded everything that could be recorded for the various indicators. All the systems included data on NWO grants, PhDs completed, articles in leading periodicals, books published by well-known publishers and prizes awarded. However, data on aspects relating to societal relevance were often far from complete; it is usually more difficult, for example, to gather such data, while too little information may be available on the underlying definitions, or it may simply not be ‘standard practice’ to report these sorts of things (media appearances, for example). The way in which these data were reported was regularly somewhat impressionistic and in the form of non-exhaustive, narrative summaries accompanied by incomplete data sets. This makes it even more complicated to assess and attribute due weight to these data than in certain other areas. The same applies in respect of viability; the very nature of this category means less information is usually available, with the exception perhaps of concerned comments about funding of law schools (often) and the need to retain the best researchers, including the risk of junior professors leaving the institution early on in their careers (sometimes).

With regard to the availability of data, it is also important to note that a number of faculties underwent varying degrees of restructuring during the period 2009-2015, and that these operations resulted in new focuses, programmes and organisational structures, thus creating an extra hurdle for the relevant committees in their assessments of the available data.

- Lastly, it is both interesting and understandable that boards of schools and research groups were keen to show that they think creatively and in depth about their faculties’ and groups’ profiles, about encouraging or winding down research groups, and about the complex interaction between ‘bottom-up’ and centralised management and the accompanying incentives. This is an inherent part of assessments, as I know from the many occasions on which I myself have sat on one or other side of the table. It regularly resulted in “dean’s speech”, as one of the foreign committee members described it; in other words, a point at which the ranking (or fears about ranking) and the ‘holding the mirror up to oneself’ cross. Too much self-reflection externally expressed can indeed make vulnerable.

It is not my task to assess those who carried out the assessments. Furthermore, I attended only parts of the visits. From a general perspective, however, I can confirm that the focus in all the visits was on frank and open dialogues, asking critical and sometimes very critical questions and continuing to probe where practice did not seem to reflect the faculty’s policy (‘You directors say that, but the message we get from the work floor is different’). An important issue identified was, again, the danger of freely expressing far-reaching ambitions: it is good to have them, but they also make schools and research groups vulnerable if they are insufficiently substantiated by data submitted. This resulted on several occasions in a recommendation certainly to retain the ambitions, but to adopt in the future a more process-driven approach, including outlining appropriate timeframes

and appreciating the need to obtain greater insight into factors influencing the ambitions and the outcomes.

In the assessments of the key publications, but often also on other occasions, two committee members consistently acted as first and second reporter, or as reporter and first referee. During the assessment visits, they were given substantial scope by their chairpersons to ask specific, subject-oriented questions and to express initial opinions. That resulted in something I see as the essence and indeed the great charm of the entire exercise: colleagues/peers holding frank and open discussions on, for example, important publications and asking why these are seen as innovative and what their relevance to society is. The initial findings, often still formulated in an impressionistic style and sometimes submitted in writing beforehand, were then refined via an exchange of arguments and translated into widely supported outcomes.

Given that the current model is not intended to produce a ranking for faculties and research groups and having established that the decentralised model represents less of a burden for the assessors, and also that faculties had the opportunity to ask specific assessors to examine specific points of concern and attention, the nine chairpersons and I concluded that the current model is essentially a good one, although in practice there are always some aspects that can be improved, certainly if more time would be available. On the other hand, there is no need to put salt on every slug, as the Dutch proverb goes. What matters is the general picture following the critical pre-reflections and question and answer sessions, and based on how the various faculties and groups accounted for their performance. Ultimately it is a matter, luckily, of 'weighing things by hand', with arguments and reasoning counting for more than rounding off discussions into numerical scores.

A total of 30 research units (faculties, centres/institutes and groups) were subject to both a qualitative and quantitative assessment, with 90 ratings or the equivalent being issued (1/excellent, 2/very good, and 3/good, including the nuances stated at the start of section 1.2.2). In terms of the individual assessment criteria, the picture was as follows: research quality: 4 x 1, 23 x 2 and 3 x 3; relevance to society: 4 x 1, 26 x 2; viability: 4 x 1, 18 x 2 and 8 x 3. Most of the variation in the ratings was at a programme rather than faculty level.

The high number of ratings of 2 could to some extent have been expected, given that the scale is not very finely meshed. After all, a rating of 'excellent' means no grounds for any substantial comments, and this is designed, at least partly, to avoid excellent ratings being subject to any further 'grade inflation'. Meanwhile the gap between 3 ('good') and 4 ('unsatisfactory') in the new 4-point scale is also substantial, with experience showing that a rating of 3 is seen both by those assessing and those being assessed as closer to 'satisfactory' than to 'good'. The overview also shows that most of the ratings of 3 related to viability, and so usually either wholly or partly to persisting uncertainties on the financial side. In some cases, committees also expressed their confidence in the future, but had too little hard data on which to base this confidence.

With regard to the numerical ratings, no unit was given a rating of 4. That, too, was not unexpected as the faculties have already made many choices in the past few decades, including transferring groups that were not functioning well to a back-burner or deciding that they should no longer operate as separate groups. Faculties and groups have done a lot in terms of quality management over the years, including self-assessments and forms of 'intervision' among researchers, and using quality management systems that were unheard of twenty years ago. The question is whether this has sometimes simply created more 'organisational stress'. This question, which is not intended to be rhetorical, is examined in more detail in section 3.4.

Lastly, and to conclude this section on the practical aspects of the assessment, the following three comments can be made:

- Before, during and after the visits, the secretaries devoted considerable efforts to sorting out the material, and this regularly resulted in 'informal coordination consultations' between the chairpersons and secretaries and to new critical questions or confirmation of previously formed

- views. These consultations obviously related not so much to key publications, but instead to research FTEs, external research grants, PhD successes and so on.
- The draft reports, which were all a co-production by the chairpersons, members and secretaries, were sent to the executive boards of the universities or directly to the faculties in mid-February, with an opportunity for fact-checking. The faculties used this opportunity to correct a few items and sometimes, via the fact-checking route, to challenge views they considered excessively critical. This would certainly seem legitimate to me, in part because factual substantiation of the findings cannot usually be separated from the findings themselves. I am aware of a few cases where this resulted in arguments and reasoning being made more specific, but no cases where this led to changes in the assessment itself or the ratings given.
 - The final reports were presented to the boards of the universities in mid-March 2017.

2. THE STATE OF THE ART OF LEGAL RESEARCH

Legal research comprises many different chambers, sub-chambers and branches, each with its own history and characteristics, and anyone comparing the self-assessments and the nine assessment reports will see many differences. Nevertheless, and if viewed from a slightly higher vantage point, one can see that the various units also have a great deal in common. This, in turn, allows me to compile a character sketch, in five facets, of legal research in the Netherlands as it currently stands.

2.1 Issue-driven and ‘wanting to matter’

The first of these five facets is the wish to be relevant and ‘to matter’. Researchers can achieve this relevance by, for instance, conducting a positive legal analysis of an issue facing legal practice. This comes closest to the classic ‘black letter law’ research. It is also possible, however, to achieve relevance by researching the legal doctrinal principles applying in a certain area of law. While the term ‘fundamental research’ is often equated with research that is not or not directly relevant, that definition would seem to me to be of little if any use in the case of legal research. Fundamental researchers, too, aim to be relevant; they may, for example and to put it briefly, try to persuade fellow researchers to start looking at a field of law or a specific legal doctrine in a fundamentally different way in the hope that practitioners or policymakers will subsequently seek to embrace that new approach. ‘Wanting to be relevant’ also implies that legal research usually starts with something seen as a problem by a researcher, a research group or by the society in general. And that, too, can come in all shapes and sizes and, in the case of legal practice, is increasingly often to be found outside the boundaries of classical normative research. Examples of such issues include the juridification or de-juridification of society, the power and powerlessness of institutions, the ‘deep state’, or the idea that the law has to take care not to become a tool of the establishment.

In all types of legal research, ‘quality’ essentially means telling the most persuasive story, with arguments based on in-depth knowledge of the material, and where the ‘jury’ comprises peers and end-users. Rather than using experiments, legal research tends instead to be based more on theories and hypotheses, asking for falsification. As a rule, this research can also be replicated, precisely because it consists of arguments that are rebuttable. As a result, the research’s validity is not based on objectively seeking to establish ‘the truth’, but instead on the idea that the better the analysis and the methodology used, the more the outcome will be accepted as coming close to ‘the truth’.

2.2 Pluralism of methods

Content is playing a leading role in research, with the method more and more often assuming the role of follower. Although it may seem obvious, for a long time the practice in legal research was to take a topic from social reality and to shape it so that it could be researched using a classical legal method. In effect, therefore, reality was adapted to the available and, for a long time, only method. These days, however, that has changed considerably, and indeed is continuing to change, in all faculties. Everywhere, complex legal and societal issues demanding attention from legal researchers and others are being discussed, with the most suitable research method being selected on a case by case basis and varying from classical legal to multidisciplinary, interdisciplinary and transdisciplinary methods. Increasingly, therefore, legal research can be said to deploy a pluralism of methods.

Over the past few years and sometimes even decades all faculties have been and are increasingly reflecting on the methodological sides of their research. At the same time, the assessment visits found that, with the exception of the established interdisciplinary institutes, a need remained for an “exhaustive reflection on methodology dilemmas”, as one of the chairpersons wrote to me, also with regard to classical legal research. This is important both internally in terms of explaining and accounting for methods used and externally with regard to other disciplines. In addition, all the faculties have assigned responsibility to one or more researchers to focus explicitly on issues of methodology alongside their ‘normal work’. However, these methodology researchers generally seem to target young researchers rather than ‘established names’. It can sometimes also be queried as to whether their work is entirely tailored to the methods needed to ensure an optimal response to ‘law-in-context issues’ or ‘multi-dimensional issues containing legal/juridical aspects’. The general idea obviously needs to be that if a researcher or research group wants to conduct well-founded multi- or interdisciplinary research, that research needs to be based on standards also regarded as appropriate in the other discipline(s), while it should be pointed out that a multi- or interdisciplinary approach to issues can give rise to new methodological insights (‘transdisciplinary’) as well, and that these insights cannot simply be approved or rejected because of how they may appear from an established methodology perspective.

Another question regularly raised in legal research in the past has been ‘whether or not a specific law has been effective’. Once more methods are available, key words such as ‘impact’ and ‘effectiveness’ automatically assume a different significance. Irrespective of that and in a somewhat different sense than meant here, all the faculties have long been involved in cross-disciplinary research, including, for example, research in the fields of philosophy of law, history of law, sociology of law, psychology of law, law and economics, criminology, law and information technology, and law and literature. These are all fields that have always operated outside the framework of classical legal research, and each had and has its own methodological issues to deal with.

Some assessment committees found that publications and centres presenting themselves as multi- or interdisciplinary did not always qualify for that in practice ‘in the true sense of the word’. This also applies in respect of PhD theses of a multi- or interdisciplinary nature: assessing such theses are ideal moments for establishing the depth of the research in combination with the extent to which it is valued by members of PhD committees coming from a background in a subject other than law. In this respect, there is still a lot to be learned by the academic world as a whole, and certainly also by the field of legal research, in the form, for example, of cross-disciplinary periodicals and full recognition of publications in such journals.

Lastly, it follows from the combination of points raised in sections 2.1 and 2.2 that legal researchers have increasingly less reason to ‘seek to avoid’ complex issues facing society and that demand a lawyer’s view. Regardless of the contents of such research, one of the practical consequences is that, in contrast to classical legal research, this research is better positioned with regard to the second flow of funding. NWO, for example, has tailored its organisation and assessment methods for humanities and social sciences, in two phases, towards research at the intersection of disciplines, primarily to

reflect the diversity of research subjects and on the assumption that this type of research will be the most innovative (see section 4.1 for more details). Whether the prospects for legal research will improve in absolute terms will obviously depend partly on developments in other disciplines as there, too, researchers are not standing still.

2.3 Variety of publication forms and fora

In my view, the traditional distinction between fundamental and applied research has for a long time dominated many discussions about the nature of academic research (also in the field of law) and the accompanying assessment-related quests. If, however, the perspective outlined above is adopted, as already happens *de facto* at all faculties, it will be far easier to accept the existence of a range of different target groups for the knowledge and insights generated by legal research and to accept that there are different ways in which this knowledge can be shared. It also follows from this that one publication form is not necessarily, let alone automatically, of less value than another. To put it differently, an annotation – which is sometimes viewed critically in other disciplines or, even worse: ‘If there are so many of them, there must be something wrong’ – can disseminate highly innovative insight and also provide the foundations for a book or longer article. Indeed, the form, length and detail of annotations in some journals can approach the level of a proper scientific article. In any event, research in the field of law includes a wide spectrum of publication forms, ranging from academic articles, professional publications, popularising texts, loose-leaf publications and books (monographs, manuals for use in education and legal practice, and edited volumes) to advisory reports for policy and research as well as practice, chronicles and blogs. In other words, there are plenty of different forms and fora available. Something commonly occurring in other disciplines, with their focus on scientific articles in ‘high-impact journals’ and the accompanying citation indexes – on which more in section 4.4 –, which sometimes has also been taken as a standard in discussions of legal research, fails to do justice to the specific nature of that research. And this not only relates to legal researchers. Literature researchers and historians, for example, also write books, as I could put it.

With regard to journals, legal research does not work with nationally recognised lists of journals and their accompanying A, B and C rankings, despite various attempts to compile such lists. For a long time, the reasoning seemed to be that all recognised disciplines used such lists, and so why not law? A national, overarching list of journals may admittedly not be possible or desirable, but it has certainly proved possible to draw up lists of journals considered prestigious in specific sub-disciplines, and these lists are used by various faculties. This applies in sub-disciplines of law that, by their very nature, operate internationally, such as philosophy of law, law and information technology, European law and my own field (international law), but also in a discipline such as law and economics, which does not tend to operate only with A, B and C-ranked journals, but also uses impact factors and citation indexes. So why should *all* researchers in these sub-disciplines not be expected to publish in a leading journal from time to time as a way of showing where they stand in an international comparative perspective? And the emphasis here is on *all* researchers: senior researchers, too, need to sweat, as one of the committee chairpersons put it. That is not the same, however, as demanding that *all* publications should be in such journals and that the rest ‘do not really count’. Adopting the latter attitude would go against the pluralism of fora that has traditionally characterised legal research and from which it derives much of its strength.

2.4 Relevance to society

Reference was made above to legal research’s ability to link scientific benefit with societal benefit. Unlike indicators for measuring the quality of each constituent part separately, this linking remains a key characteristic of much legal research, even though it may be problematic from a perspective of

the same efforts to measure quality in line with the existing protocols. The increasing recognition of societal relevance in the SEP, as also adopted in the Discipline Protocol, shows that, in some ways, legal research is ahead of its time, also by its acceptance in principle that linking research to ‘practice’ can result in the co-creation of knowledge, that high-quality theoretical research also exists in part due to its shorter or longer-term relevance and that, seen more broadly, legal research does not take place in an ivory tower, but is instead a domain where different worlds and perspectives meet and cross-fertilise. It was no different back in the days of Hugo de Groot.

As indicated earlier, these various aspects combine to result in specific publication channels being chosen for specific purposes, but also, for example, in *contract research*. For all faculties, contract research is both a way of utilising existing knowledge and gaining new knowledge and also a welcome and sometimes vital source of additional income, which is needed to enable all the aspirations to be achieved. The interaction between the research group’s academic profile and contract research is a recurring issue of attention and concern. Ideally, demand for contract research will and usually does arise from the academic status of the research group (and the external awareness of this status). Contract research can certainly also be of a fundamental nature, with the challenge being to distinguish between ‘fundamental’ (i.e. depth of content, and reflecting on principles etc.) and ‘independent’. But there are also concerns largely relating to the second aspect: i.e. can a group permit itself financially to refuse to conduct certain research? What if the reliance on specific parties providing funding becomes too large? And to what extent can and should researchers accept restrictions on their freedom to publish? However, although widely ventilated within the faculties, these concerns are not standing in the way of law faculties in the Netherlands performing contract research on a large scale. And whatever principles one may have, it is sometimes hard ‘to look a gift horse in the mouth’, and personal integrity is then the only standard remaining (see section 6.4 for more on integrity).

Another issue arising in some faculties relates to the fact that topical *debates in society* are regularly conducted without any knowledge of the legal issues involved. Although law is not always relevant to certain issues affecting society, in those cases where it is relevant, it cannot do any harm for legal researchers to let their voices be heard, certainly in an age where in-depth analysis increasingly has to compete with opinion. Law can, after all, provide a certain ‘infrastructural and institutional foundation’ in times of turbulence. Just think, for example, of the judiciary’s response to the travel restrictions imposed on certain countries’ nationals by US President Trump, or South Africa, where courts have thwarted President Zuma on various occasions. Interventions by legal researchers in public debates can also help ensuring that debates do not keep on having to be restarted from the beginning (‘law as stilled morality’, however fluid and subject to change it may be). And this is not simply a matter of radically countering ‘alternative facts’ and opinions presented, intentionally or otherwise, as facts, but also of revealing differences of interpretation, including allowing scope for different views. In essence, this is about helping to monitor and actively promote the quality of societal debate, bearing in mind the greater good of a well-informed democracy. Another not unimportant point is that the law cannot speak for itself and so needs spokespersons who are willing to stand up. They can also make their contributions by highlighting fundamental principles of law, such as the right to be heard, the principle of innocent until proven guilty, the separation of powers, and the precautionary principle. Although these principles are not prerogatives of legal research, they usually arise from legal research and are regularly examined by legal researchers for their current theoretical and societal significance. As a well-known saying among lawyers goes, ‘You might well be learned’, to which my response would be: ‘But let it be seen (or seen more)’. Because there really is a lot at stake.

2.5 Internationalisation

Only a few decades ago many people, both within and outside law faculties, thought that ‘internationalisation’ went against the very essence of a world in which the law studied was first and foremost national law. And if an international issue arose, the obvious response was to seek recourse to the method of comparative law. These days, however, and although this method is still widely used, it now tends to be more and more accompanied by consideration of issues and themes/concepts of European or global significance, also outside the sub-disciplines that have always operated in this way. Here, too, one can see wide-ranging participation in international debates, with PhD candidates being encouraged to look beyond national borders, and university staff coming from more diverse backgrounds. For many faculties, groups and individual academics and other law professionals, internationalisation has become second nature, alongside the attention they continue to devote to their national law. At the same time, the extent and shape of internationalisation very much depends on the separate sub-disciplines, while it is also vital to distinguish genuine internationalisation from simply publishing in English, French or German, given that publications in these languages may relate purely to matters of national law. Some faculties were also found to have set up committees to promote internationalisation or its equivalent, while others have since moved on and yet others have re-established committees to map out the route ahead. Like many aspects of this report, this, too, reflects the evolution of legal research, with the developments in this respect combining to bring the discipline to the *state of the art* outlined in this part of the report.

3. ORGANISING LEGAL RESEARCH

Before turning to ‘measuring’ the qualities referred to above and the indicators that are available for this purpose, one subject still deserves special attention: how is legal research being organised? Here again, the faculties show more similarities than differences.

3.1 Programmes and group sizes

Legal research has traditionally shown – and to a large extent still does show – an individual orientation. For a long time, this has influenced discussions of programmes and programming in legal research. It has been argued that the nature of legal research resists being programmed as, for example, there always has to be a relationship to rapidly changing societal issues and realities. In my opinion, this argument, or counter-argument, is based on confusion regarding the term ‘programme’, as if this means that everything has to be ‘programmed’ in advance. This having been said, the assessments have shown that legal research programmes are often quite loosely connected and that groups of researchers often do not focus on one specific subject, let alone one specific research question, in contrast to what is common practice in some other disciplines. Programming is often tailored to the researchers, as it were, instead of the other way around, and programme leaders and senior researchers are then expected to assume a leading role and to work with the people and the means at their disposal.

At the same time, and despite all qualifications, programme titles are often a good indication of the contents of the field of activity. Meanwhile there is a broad consensus, also in legal research, on the value of and need for some research programming and the ensuing formation of research groups. Some keywords regarding contents and organisation in this respect are: the opportunity to formulate more extensive, long-term research questions that exceed the capacities of one

researcher; the formation of research groups around certain subjects, whether for clearly defined periods or otherwise, and often also aiming to promote programme coherence and external visibility; the combining of disciplines for multi- and interdisciplinary work; organising combinations of senior and junior researchers; and, last but not least, the increasing of opportunities for funding from organisations such as the Netherlands Organisation for Scientific Research (NWO) and the European Research Council (ERC) (for details, see section 4.1).

During the site visits, reservations were expressed regarding the requirement for research units to comprise at least 10 research FTEs (amounting to approx. 25 senior researchers, again excluding PhD candidates and postdocs), which some felt to be too large from a perspective of effective decision-making, and of working together on a project, providing an inspiring research climate, and ensuring that the management of the group itself is still able to make substantive contributions to the research. There is no such thing, however, as an ideal group size, and nor is there any clear correlation between group size and high ratings: some small groups have modest ratings, while some larger groups were rated 'very good' or 'excellent'.

3.2 Position and role of PhD candidates

All faculties widely support the view that PhD candidates make a substantial contribution to the refreshing of research blood and to innovating research lines. Writing a PhD thesis is also increasingly seen as a regular third or fourth educational phase, and aimed both at personal development (the writing itself, working in a team, developing an investigative attitude) and at preparing for a job outside the world of legal research. For various faculties, this is a reason to offer combined research and education contracts so as to provide employees with a more all-round education for the job market.

This being said, plenty of questions still remain as regards PhD candidates. I will mention a few that play a role, to varying extents, at various faculties:

- How can exchanges between faculties be encouraged, whereas faculties are, as a rule, hesitant to let their best students go (and then often only if forced to do so)?
- Will sufficient Dutch-speaking PhD candidates come or stay, with a view to promotions to higher positions and to subjects required to be taught to those students wanting to practise Dutch law (*civiel effect*)?
- Are PhD candidates, who as a rule are starting researchers, sufficiently supervised at the outset to ensure they are working on feasible subjects?
- How to ensure that the primary focus remains on the research contents and does not shift to the method or to financial remuneration for the finalisation of PhDs, or even to a dependence on such finance?
- How rigid should supervision be, and how much initiative and autonomy may be expected of PhD candidates, also in view of later activities within or outside the field of legal research?
- Are there consequences if supervision is inadequate?
- Is there sufficient confidence in the functioning and independence of the deans for PhD candidates (or their equivalent) in the event of complaints about PhD candidates' supervisors?
- Are the faculties prepared not only to accept but also encourage candidates to obtain their doctoral degree by publishing articles, e.g. with a view to increasing the chances of an NWO-Veni grant?
- What steps are being taken to ascertain that, in cases of a co-authorship, the most creative parts are written by the PhD candidates?
- What are the consequences for authorship of increased multi- and inter-disciplinarity? Will traditions from other disciplines (e.g. listing professors who devised a subject as co-authors) become more frequently used in legal research?

- If someone obtains their PhD degree by publishing articles, how closely related should the articles be, and should an introduction and a conclusion be provided, or is this view modelled too much on the traditional idea of a book?
- What exactly is the role and significance of PhD committees if articles have already been published as *refereed* articles?
- How can external PhD candidates' involvement in the academic culture that already exists for regular PhD candidates be improved, including quality assurance through intermediate steps, so that the quality assessment by PhD committees ultimately proceeds smoothly?
- How can completion times and output be improved, without PhD candidates being regarded as a 'production factor' and with realistic account being taken of the substantive, organisational and personal challenges a PhD candidate often faces when writing a PhD thesis?

3.3 Instruments for steering and control

In the discussions on the programming of legal research, the keyword during the site visits often was *bottom-up*, given that researchers and research leaders are best able to judge what is 'in the air' in their specialist areas and will consequently also identify with their subjects ('ownership'). Researchers are regularly averse to being pawns on an administrative chessboard, which also asks for a corresponding management style: allowing space; not be tied down too much or burdened with too many internal protocols; allowing to focus primarily on the creative, innovative and imaginative aspects of their work, and avoiding 'management by distrust', especially *ex ante*. Essentially, this involves the traditional academic freedom: what do I tackle as a researcher (and a research group) and how do I go about it? It is my impression that this freedom still exists to a considerable degree, especially for senior researchers, and that the aspects restricting it are the demands on time that good education imposes and the bureaucratic procedures that have encroached on the universities over time rather than substantive choices imposed by the faculties and universities from above. This still offers opportunities for 'curiosity-driven' research (as NWO generally phrases this). In my opinion, anyone who has the talent but not the time for this should prioritise rather than complain.

It is important, however, to closely monitor – both faculty and university-wide – the 'indirect steering of research', both because of the dependence on external funding and because funding bodies such as ERC and NWO frequently have their own content agendas. And even though programmes funded by those organisations are generally based on consultations with the potentially relevant sectors, these agendas are not advantageous for certain legal sub-disciplines or even legal research as a whole.

With the bottom-up approach as the starting point, all the faculties still have ample scope to steer and adjust their research. I will mention just a few 'instruments', some of which came up during discussions and in the nine reports:

- Releasing funds to recruit PhD candidates and postdocs for specific, centrally selected subjects.
- Enabling all researchers, both young and old, to go abroad for a while as a regular researcher, or as part of educational leave or a sabbatical, including for those who are not former (vice-) deans etc.
- Making small scholarships available for innovation ('sowing money'), which can be spent freely, without any need for accountability.
- Establishing a faculty innovation fund to enable researchers to work on larger proposals on an ongoing basis.
- Appointing experts in external fundraising and with specific knowledge of the legal domain and its societal context.
- Organising faculty 'dating sessions' to find out which researchers are interested in cooperating on particular subjects.

- Making ‘credit appointments’ for professors (who are promising and ‘made of the right stuff’, but whose past performance is still quite modest).
- Supervising young PhD supervisors.
- Appointing research professors.
- Entering into work and performance agreements with research groups performing less well, including ‘sunset clauses’, and more quickly phasing out substandard research, also to allow for more promising initiatives.

There are many more instruments, each with their pros and cons. With a few exceptions, selecting one instrument will have financial consequences for other activities, within or outside the research domain. Between having a dream and achieving one often finds financial constraints.

The essence of combining the bottom-up approach with top-down management and control is to innovate without detracting from proven attainments. However, this is often not easy. During the earliest debates on the internationalisation of research, for example, researchers with an extensive track record in national legal research sometimes felt the ground disappearing from beneath their feet. These days, however, there seems to be more of a balance, with opportunities for variety, providing the quality of the work carried out meets the standards. What ultimately matters in the interaction between the bottom-up approach and the managing, controlling and adjusting of activities on the one hand and the various instruments available on the other hand is, once again, the design and management of the integral chain: ranging from carefully formulated research ambitions to the drafting and implementing of a financial model for the faculty and a staff policy, including incentives, both of which are realistically tailored to those same research (and educational) ambitions.

3.4 Interaction between contents and organisation

Both in conversations and in their reports, several of the nine committees expressed the view that, to put it mildly, the faculty organisational structures are not always very transparent to outsiders. Apart from the various names (centres, departments, programme groups and their umbrella organisations), this lack of transparency includes the relationship between the different entities (e.g. the role of the *graduate school* in relation to centres and departments); the question of whether interfaculty institutes do have an autonomous position; the position and authority of the vice deans of research; the lines between the faculties and the executive boards of the universities. And on another level: researchers belonging to more than one research group and department, with the accompanying questions on deployability and human resources policy (e.g. who will conduct the performance reviews?). The committees frequently asked the boards whether they would organise their faculties in the same way if they had a chance to start all over again. Some of them had ‘a hard time’ in explaining why they would. And not to forget: each faculty clearly has its own history, which also places a strain on organisational forms. The exchange of views between the committees and the boards on such matters frequently resulted in the committee chairperson stating that ‘it is okay as long as the organisational structures are clear internally’.

The principal question remains as to whether substantive questions and ambitions are the guiding principles in reforming and maintaining organisational structures, as well as its mirror image: the fact that each change consumes time and energy. The latter in turn raises the question of how – other than in times of forced reorganisations – to avoid turning things upside down, without the changes being instrumental in achieving the research (and educational) ambitions. In addition, measures emanating from the government or the central university level sometimes require other administrative measures than those that the faculties would have come up with themselves, while they are also part of this larger entity. The same applies to research groups within faculties. They, too, are part of an administrative context. Furthermore, education and training for legal practice, especially with a view to *civiel effect*, may require different structures from those required in legal

research. Further to this, the growing emphasis on multi- and interdisciplinary research seems to demand a more thorough investigation of faculty organisational structures than is currently often the case.

This leads to the more fundamental question of the relevance of research policy in general, with the central question being: does it do what it claims to do and what would be the state of legal research without controls, adjustments and interventions? It is impossible to answer this question, especially in a general sense. I am convinced though that the successive assessments of legal research and the ensuing policy discussions and steps have greatly benefited the discipline, with all the incentives they have provided for self-reflection etc. The fact that there is no easy answer to this question is also no reason not to pose it. Time and again the plain objective question has to be raised as to the added value of policy steps, and whether it will make a positive contribution (i.e. improve quality) or a negative contribution (i.e. increase bureaucracy) to the protocolisation of research.

4. QUALITY AND MEASURING QUALITY: REFLECTIONS ON VARIOUS INDICATORS

In part 1 of this report, I indicated that there are many possible indicators of quality, also in the case of legal research. One should keep in mind that an ‘indicator’ is called an ‘indicator’ for good reason, i.e. it is an indication of quality rather than evidence that requires no further human action. And although the accuracy in the faculties’ use of the existing indicators may have impeded some parts of the current assessment (see subsection 1.2.4), this does not alter the fact that they have made significant progress in this field in the past few decades. A broad consensus has been reached on several indicators of research quality, relevance to society and viability, i.e. the current key assessment themes, which will most likely also be the guiding principles after 2021, albeit possibly with new accents. I will discuss a number of these indicators and make some modest suggestions for improvements. However, these suggestions must again be seen in the context of the preliminary management question as to which route the faculties, individually or collectively, will choose for future assessments.

4.1 External research grants

The awarding of an external research grant by NWO and, for example, the ERC is held by all faculties in high regard. In respect of the ERC it should be noted that some of the previously available national funds have been shifted in recent years to the EU, partly based on a substantive agenda in the field of law and regulation, among other things. Some faculties realised this in good time and were able to achieve a good score, particularly with regard to FP7/*Horizon 2020* funding. It is clear, however, that legal research could be more ambitious about competing for ERC grants, even though many initiatives have admittedly already been taken. In the case of both ERC and NWO, there will always be discussions on the procedures and on issues such as the relationship between the time that has to be invested and the chances of success. However, all faculties are proud to report, for example, the awarding of Veni, Vidi and Vici grants, and they invariably present such awards as a mark of external recognition and, therefore, of quality. This causal relationship may not be as self-evident as it seems; however, here, if anywhere, the hard facts can speak for themselves. According to an overview I received from NWO, those facts are that, between 2009 and 2016, 31 Veni, 11 Vidi and 4 Vici grants were awarded to legal researchers. Compared to other social sciences and humanities, legal research scores slightly below average for the applications/awards rate, but it is not as bad

as some may think. In the relevant NWO domain, comprising social sciences and humanities, the success rates for the Veni grants for the years 2009-2016 were 12%, compared to 10% for legal research, while the rates for Vidi grants in both categories were 11%, and for the Vici grants were 13% and 8% respectively. For the record (methodologically speaking), these figures are purely indicative. I have not, for example, taken into consideration the absolute numbers of applicants, nor the possibility of disciplines (and faculties/groups within disciplines) submitting everything, the good with the bad. Although a closer investigation of this will not produce different numbers, it will provide a more balanced insight.

In contrast to this relatively positive image, the Spinoza grant, for example, which was introduced in 1995, has only once (out of a total of 81 times) been awarded to a legal researcher, Deirdre Curtin (2007). While it is true that, as regards the Spinoza grant, social sciences and humanities do not score well, having received roughly 20% of the total number of awards, language and literature researchers have been the welcome recipients of these grants on five occasions. Although no quantitative data are yet available, it is also common knowledge that legal research has profited little from or contributed little to the Top Sector Research, Gravity Grants or NWO programmes such as Responsible Innovation. Here I will draw a line, given that a more extensive investigation of this aspect would exceed the plan and framework of this report. I can conclude, however, that those faculties that have so far performed less well in the ERC and NWO field in comparison with their 'competing colleagues' have invariably expressed a desire to make progress in this field by, for example, devising internal incentives and screening procedures through a series of channels designed to detect proposals' strengths and weaknesses.

4.2 Third, fourth and fifth flows of research funding

In my view the above also applies, *mutatis mutandis*, to large third flow of funds projects that have been contracted out to specific research groups because of their expertise, and to large amounts of funding secured under the banner of the fourth flow of funds (external research funds that are not related to specific research questions, but to a wider research field or theme) or the fifth flow of funds (where research groups 'tell' the sponsor what the research subject should be). In the current assessment, frequent references were made to large third flow of funds projects, which in some cases, in conformity with the protocols, also resulted in the outcomes being listed as key publications. Various examples of funding via the fourth flow of funds were also presented, however not of the fifth flow of funds. Following the example of other disciplines, this latter option may, however, hold prospects for some legal research groups. In my view, and providing they are accompanied by a thorough explanation of the relationship between expertise and contracting-out, these achievements can unreservedly be regarded as a sign of quality, despite any objections conceivable in theory. I would like to point out, however, that potential developments such as these should not automatically be included in protocols, let alone be included as tasks to be fulfilled in bilateral agreements between executive boards of universities and faculties. The reason I mention them here is purely to keep the door open for the future.

4.3 External peer reviews and assessments by editors

An external *peer review* of publications, with the *double-blind* principle often being regarded as the absolute, and again with the comment that the SEP does not use these terms and only mentions '*refereed*' publications, is in principle a sign of proven quality. 'In principle', as a *peer review* also has its disadvantages and downsides: depending on the *peers* involved, very innovative papers may either be misunderstood or dismissed as being too radical. Sometimes *peers* are not yet acquainted with the above-mentioned pluralism of methods and, more in general, with the state of Dutch legal research. Some international assessors said that the Netherlands was running ahead (or too far ahead), or

words to this effect. Comments I recorded included ‘Trend-setting’, ‘You are real trailblazers’, ‘Your legal scholarship belongs to the top’, ‘The Dutch approach is admirable’ or, alternatively, ‘You are way too Calvinistic’, which was said when draft assessments had to be formulated in a confidential final discussion. This type of criticism is also heard when legal researchers submit proposals to the ERC or NWO.

Furthermore, it is not always clear whether journals and editors who claim to use a *double-blind peer review*, for example, actually adhere, or adhere fully, to this principle. And there are more arguments, including the absurdity of the reverse: i.e. anything that has not been subjected to *peer review* cannot be good. How should we, for example, interpret the quality of an article that was judged ‘only’ by a board of editors, but published in a journal known for rejecting (for content reasons) high numbers of articles submitted? It would seem sensible not to rank the concept of external *peer review* higher, and certainly not automatically higher, than an evaluation by highly critical boards of editors (these editors, too, are also *peers*), assisted by experts in sub-disciplines on an *ad hoc* basis or otherwise. In addition, authors and reviewers of nationally oriented research will often know each other, particularly also if they work within national research schools. As indicated in section 1.2.4, there is scope to specify definitions more precisely and to monitor their use in practice. And if anywhere, the focus here should also be on the whole chain: from reaching good agreements on definitions for national or local/faculty use to correctly recording data and monitoring how these data are fed into ‘the systems’, so that ultimately the assessments are not dependent on researchers’ own judgement.

4.4 Impact factors and citation indexes

Some faculties make use of impact factors and citation indexes for specific research units. To me, it would seem that the trend is to let this continue in the sub-disciplines currently using such a measuring method, but not to extend it to other sub-disciplines, let alone to apply it faculty-wide. In my opinion, this is the best approach, also in view of the discussions being held in other disciplines on how these indicators can become polluted: not every citation or every form of impact is a mark of academic quality or high relevance to society, although someone who has been cited frequently will have less explaining to do than someone whose work has never been cited; the same applies to impact. Other important considerations are that many legal journals are not included in the large databases that are used to register impact and citations, while citations in books, for example, do not count. The fact that some authors never cite from the work of certain colleagues/authors also puts the importance of citation indexes and impact factors into perspective.

4.5 Dealing with concurrence of academic research and practical relevance

A complicating factor in measuring quality in the legal field is the frequent concurrence of academic research and the relevance (or envisaged relevance) to practice. Something that, in my view, is a commendable characteristic of the contents of much legal research poses problems from the perspective of quality measurement and the precision of the underlying indicators, both with regard to the definitions used and the recording of the output. An example may illustrate this: although the *NJB* (a Dutch law journal) distinguishes between ‘academic research’ and ‘practice’, articles in the ‘academic research’ section are not always strict in a methodological sense, while the same holds true for ‘articles from practice’, the nature of which is can quite often of an empirical nature. In other words: certain aspects of indicators start to overlap here, and, again, this is a measuring issue rather than a question of the articles not being good or not deserving a place in the range of qualitatively good research.

Interestingly, the Discipline Protocol presumes the sub-criteria for the quality of publications to be sufficiently well known. Yet it is important to regularly hold these up to the light, whether

as part of an assessment or otherwise, and to actively remind everyone of the core characteristics of quality: innovative, broadening or deepening one's knowledge, an evident addition to the *body of knowledge*, a clearly formulated methodology (of any form), with all of this being subject to supervision that extends beyond researchers' judgements of the quality of their own work.

4.6 Relevant positions and significant invitations

Another set of quality indicators relates to KNAW memberships, honorary doctorates, guest professorships and, in a way, prestigious academic invitations. However, the last of these indicators is, in my view, prone to inflation to a greater extent and in another way than the other above-mentioned 'marks of external appreciation'. Many organisers do resort to these words, knowing that presenting a *key note* sounds prestigious. Yet not all *key notes* are of equal value. If they continue to count as a 'quality indicator without any substantive assessment', my suggestion would be to count them in full only if the invited speaker is able to submit a formal and reasoned invitation. Again, this is a matter of weighing 'a little more bureaucracy' against an 'objectification of the weighing', with the additional advantage of this being that such a step would do justice to those who regularly receive such prestigious invitations indeed.

4.7 Membership of editorial boards

In my view, care also has to be exercised when using editorial board memberships as an indicator of quality. It is not always the best researchers who are invited to join such boards, or it may be felt that 'They have already had their turn'. The editorial board may also have a limited pool to choose from, as a result of which everyone automatically 'takes their turn'. This is not a plea for deleting this as an indicator, but instead a call for such memberships to be specified in more detail. This could be achieved by designing a form to include criteria such as national-international, rejection rate, open access and a job description of the editorial board member, as this would allow editorial board memberships to be categorised. In this way, quantitative data will ultimately feed the qualitative judgement in a more reliable way than is currently often the case. However, it is *once again* up to the research groups and the boards to weigh the substantive pros against the management and administrative cons and to ask themselves how much weight they ultimately wish to attach to this indicator. Leaving things as they are is also an option.

4.8 Knowledge for policymaking

With respect to other quality indicators, specifically, for example, with regard to relevance to society, it is possible to devise sub-indicators, such as 'knowledge for policy', which could include items such as traceable contributions to legislation, to advocate-generals' opinions, to government policy advisory bodies and committees, as well as to the private sector. This would also reflect the university-wide trends towards greater valorisation. Although dividing indicators in this way would create greater transparency and thus more opportunities for external verifiability, it would again firstly require a weighing-up of the management and administration costs involved.

4.9 In conclusion

Each indicator has its own narrative. Content is leading and much can be captured in indicators. However, these will remain indicators in the narrow sense of the word. Moreover, it is hard to capture some matters in verifiable indicators, such as the question of whether a research group 'vibrates' and whether there is an open interaction between management and the 'work-floor'.

Another question is how this same work-floor views controls and adjustments based on indicators. The issue of how faculties record academic output in scoring or ranking systems or their equivalents arose in discussions at various law schools; it is interesting that some researchers reported seeing such systems as a further bureaucratisation of the research work. If designed correctly, however, these systems should reflect the faculties' priorities as regards content. Nevertheless, it would seem a good idea for faculties to hold up their systems to the light again after this assessment round. Some faculties are considering, for example, taking steps to further encourage multi- and interdisciplinary research (or have already started doing this) by including this type of research in their rating systems. The existing systems are not disregarded, but can instead be re-examined and possibly refined to reflect the faculties' own ambitions and the insights acquired through the assessments. This applies throughout the chain, which this time ranges from policy to organisation, incentives, research efforts, and results, structured transparently and with quality indicators that are manageable in practice, also and not least for all those researchers *not* dealing with faculty policy issues on a daily basis.

5. BENCHMARKS

In the Discipline Protocol, I was invited "to express my opinion on the way in which benchmarking has taken place and may take place in the future" (p. 10). The Protocol notes that benchmarking constitutes a new element in the SEP and that "in this transitional phase the law faculties (and the executive boards of universities) [seem to] have different views of benchmarking". Moreover, there is "still little experience with benchmarking within the discipline" (p. 9).

The SEP refers to benchmarking as a tool for reflecting on the position of research units in comparison to other research units and in society, and for exploring opportunities for the future. It adds that, for an academic institution, this means two things: "A form of benchmarking by looking at the performance of comparable units elsewhere, and an analysis of its own mission in relation to developments in the Dutch and European policy context" (table D4). I have included the context in which the SEP mentions benchmarking as this refers clearly to the position of research units in both research and society, and to the groups' own ambitions ('mission').

The core of benchmarking is "looking at the performance of comparable units elsewhere" (*idem*) and this may be interpreted both in a competitive sense and in terms of "being able to learn from one another". The SEP and the Discipline Protocol do not express a preference for one meaning or the other. The Discipline Protocol distinguishes between comprehensive benchmarking and benchmarking on "dimensions" such as academic quality, appeal, ambition or publication profile. Aggregate levels may also differ, providing the units are comparable. The Discipline Protocol also states that "some form of benchmarking – in conformity with the SEP – will preferably be part of the self-assessment" (p. 10). Faculties are free to make their own choices in this respect, albeit subject to one restriction: "No benchmarking of productivity among the sister faculties" (p. 9).

5.1 Actual use of benchmarks during current assessment

All the faculties used the term 'benchmarking' primarily in the sense of a 'standard' for measuring the quality and quantity of their own performance. The starting point is often 'benchmarking one's own ambition', including things such as numbers of *peer reviewed* articles (per FTE or otherwise), demonstrable relevance to society, numbers of PhDs completed and external funding. This part of my report relates to that other form of benchmarking: whether faculties or groups included specific research units from which they think they may learn something or with which they wish to be compared.

Some faculties have decided against benchmarking in the latter sense, with or without giving reasons for their decision. These include *Groningen*, *Maastricht*, *Nijmegen* and *Tilburg*. Other faculties have decided against faculty-wide benchmarking and have left it up to the research programmes to choose their own benchmarking, where ‘left up to’ can sometimes be interpreted as ‘advice’ and sometimes as ‘urgent request/order’. These include *Leiden*, *Utrecht* and *both Amsterdam faculties*. Only the law faculty of *Erasmus University Rotterdam* has organised a wide-ranging “strategic benchmark”. This was with the Law Faculty of the University of Copenhagen and resulted in a “concise qualitative benchmark study” (self-assessment, p. 49), a 6-page summary of which is included in the Rotterdam self-assessment report.

Apart from the self-assessments, discussions at some faculties revealed that they had decided against benchmarking as the phenomenon was felt to be unsuitable for legal research and more appropriate for disciplines where researchers “are searching for the same thing”, such as a drug for global illnesses or as yet unknown planets. In that same line of thought, legal research shows so many dissimilarities, even if the national character of a substantial part of the research is disregarded, that setting a benchmark would not result in any concrete insights whatsoever. From this perspective, benchmarking refers to ‘the desire to be the best’ and not so much to learning from one another, whether with regard to specific aspects or otherwise.

Having examined how the faculties have actually dealt with benchmarking in their reports, one can see that faculties they themselves and their research groups have alternated between the two types of benchmarking and sometimes used both (‘learning’ and ‘competing’). The points in question are often aspects where they can learn from other research units, but indicators/reference points are equally often listed with the obvious goal of seeking to be the best by comparison. In my view, neither of these approaches is wrong, with the caveat that if the result of the comparison is that the relevant institute/centre/programme *is* in fact better than ‘the competitor(s)’, this result should as a rule be put into perspective, given that it is based on the perception of research units themselves and often combined with a representation of the competitor’s characteristics that is sketchy and, though generally probably correct, hard to check.

Furthermore, many research units that have ventured into benchmarking have formulated their ambitions in quite open-ended terms, often identifying the most relevant indicators, but frequently without specifying them and thus without going down to the ‘level of evidence’. My use of the term ‘ventured’ is meant as a compliment: although their benchmarking is often not very concrete and verifiable, these units have at least attempted to undertake a benchmarking exercise. Apart from some exceptions for established research groups, the lack of concreteness and verifiability was perhaps unavoidable at this stage. In that sense, the lack of experience with benchmarking and the underlying reluctance to start doing it, as referred to by the Discipline Protocol, was found to be correct.

5.2 Benchmarking: looking ahead

Benchmarking has its origins in the business world (Japan), with its aim being to copy from other manufacturers everything that could improve a business’s own products and so acquire a bigger market share. This is where ‘learning from’ and ‘competition’ coincide. Much of the terminology used in the self-assessments would not be out of place in the business world. In my view, the universities could have learned something some decades ago from the business world on matters such as management (and reflecting on models of management), efficient time management and accountability (in the case of large companies, primarily to the shareholders and to other stakeholders; in the case of universities, primarily to the taxpayer). Having said this, there are many reasons for universities to continue emphasising their own nature: PhD candidates should not become ‘production factors’, as I wrote earlier (section 3.2), but should instead be seen first and foremost as young people who at some time in the future ‘will call the shots’ in whatever roles they take on, and who should be

enabled to continue learning and developing a sense of values, team spirit, an investigative attitude and societal daring, while avoiding a narrow outlook and becoming a super specialist in a topic for which demand may be low. Not only does a PhD programme prepare for the awarding of the degree, but it should also form a person's backbone, as a South African colleague once remarked. Although performance agreements are useful, they sometimes also have a perverting effect. Key indicators are convenient, but they can easily start leading a life of their own. Rankings? These are, or in any case seem to be, an inevitable phenomenon in the call for transparency and accountability by the society ('management by distrust'), but they are often based on selectively handled content – e.g. articles in *refereed journals*: these are crucial, but they do not tell the whole story – and use methods that do not meet the critical test of thorough empirical academic research. And it is precisely those 'nuances' that tend to get lost in the internal and external use of rankings.

Sound comprehensive benchmarking with any leading national, European or global legal research unit is, as a rule, not feasible. If, however, benchmarking were to focus on dimensions such as 'research quality', 'relevance to society' and 'viability' (and possibly numbers of research FTEs, for example, or available funds), that could in my view very well be called comprehensive benchmarking, given that it goes to the core of the matter. Surely it cannot be the intention to use minor differences in order to avoid comparison. This, of course, all fits in with the 'competition model', while the preliminary question remains whether faculties/groups actually want this model, and, if so, what exactly their aim is: do they 'want to be the best' or, for example, to expand their reputation and thus their opportunities to attract funding (a combination frequently made in the self-assessments)? One does not exclude the other.

This is all less complex in the other model, which relates to 'learning focused on specific aspects'. Based on knowledge from research units' networks, it is often relatively easy to determine that one partner excels in *refereed* articles, the other one in acquiring external funding and the third one in societal impact. The idea of this type of benchmarking is to strengthen the research unit's own 'body' and to embellish it with the limbs that support the desired self-image ('ambitions'). In my view, groups thinking along these lines must be able to explain, also to non-peers, why they choose to borrow from a particular partner. The self-assessments often cite names of other institutions, particularly foreign institutions, not only in the sections specifically relating to benchmarking and often without any specific explanation of their reasons for doing so. Or, to put it another way: not everything done in Cambridge, Harvard or Oxford is good. Anyone wishing to focus on comparisons or learning on specific aspects will have to dig deeper than often has been done so far, and cannot ignore institutional factors such as group size, funding and research time. And this must be done with the precision that is characteristic of much legal research if it is to be more than benchmarking 'off the cuff'. Here, too, the point made earlier holds true: a unit first has to decide whether it wants to benchmark and sees the point of this. But *if* it does, this benchmarking should be balanced, transparent and verifiable.

6. CONCLUDING OBSERVATIONS

6.1 Three consecutive research assessments

In 1996, the Van Gerven Committee was instructed to assess "research quality", "research productivity", "research and, where possible, its societal/technological relevance" and "the scientific perspective". It had to do this "based on largely self-conceived criteria (which were thus open to discussion)" (p. 5, final report). These criteria were discussed exploratively and resulted in many

insights, which later filtered through into the debates on the character of legal research, legal research itself and the next research assessments.

Six years later, the Ten Kate Committee was instructed to assess “research quality”, “research productivity”, “relevance” and “vitality/sustainability” (p. 15, final report) and commented that, in line with proposals made by the Van Gerven Committee, the focus was less on quantitative output data than before. Furthermore, it concluded that numerical comparisons were not reliable and that programming was not an impossibility in the case of legal research. One can say that some matters are of all time, while others refer to something that at the time was apparently still a subject for debate.

Seven years later, the Koers Committee referred to legal research as a “discipline in transition”, “if only to prevent [outsiders] from judging this report and its conclusions against the background of outdated ideas on and images of legal research” (p. 30, final report). The subjects to be assessed were “quality”, “productivity”, “relevance” and “vitality/attainability”. The committee concluded that the self-assessments still strongly emphasised quantitative information, but also that legal research had internationalised substantially, that increased importance was being attached to multi- and interdisciplinary research and that “the methodology of legal research was no longer implicitly obvious” (p. 33). It also pointed out that, despite all the programming, “in the end, the individual researcher’s quality determines the quality of his or her research”, and posed the question as to what enables an individual researcher to achieve the best results: “Research by yourself in your own space, or interaction with other researchers?” (p. 34).

Together, these short descriptions, which do no justice at all to the depth of the successive reports, present a portrait of twenty years of assessing legal research. They put things into perspective and show how well the present assessment fits into an evolutionary trend: again the discipline has made much progress, and again a lot still remains to be done.

6.2 Similarities and differences between the faculties

In part 2 of this report, I discussed ‘the state of the art of legal research’, focusing on similarities between the faculties. It is obvious that substantial differences exist as well, but first and foremost these relate to matters identified by the nine committees and are thus not differences *between* faculties, but *within* faculties.

As regards the differences *between* faculties, I would like to point out two things. Firstly, some faculties recognise more than others that high-quality, classical legal research as a value in itself may be further enriched by the posing of questions from a legal theory, legal philosophy and legal history perspective, which will, in turn, result in other types of research questions and thus, as a rule, to progressive innovation. This is distinct from the research within these sub-disciplines themselves, while the two approaches combined can contribute to the academic character of legal research, without this implying that the latter is able to break free from legal practice or ‘social reality’ in a wider sense. What is, in practice, sometimes referred to as *freischwebende Intelligenz* (i.e. ‘academic’ in a pejorative sense) may conversely serve to reinforce the distinctive character of legal research at universities. The increased emphasis on empirical, multi- or interdisciplinary research also contributes to the academic standing of legal research. This requires, however, that the legal theorists, legal philosophers and legal historians – as well as the legal sociologists, who already tend by definition to do this – and other ‘meta-lawyers’ do relate their insights to one or more specific legal fields and to do this with sufficient knowledge of the applicable law in the relevant field(s), to be able to speak with authority and, partly thanks to this, to connect with lawyers in the relevant domain. To the degree to which they succeed in this, the adjective ‘meta’ is then no longer appropriate, because they become an integral part of positive law (*‘juridica’*).

Secondly, the faculties’ profiles vary considerably. Most of these differences, however, do not relate to, for instance, the field of methodology, but instead to substantive themes, with centres

and institutes functioning as signboards of these themes. Anyone taking a closer look will see that although many faculties show a substantial resemblance, partly because of the educational obligations imposed on them by *civiel effect*, they have also created a profusion of profiles, based on the themes in which they excel or wish to excel. The experts know that although faculty X pays attention to subject Y, faculty Z has a *real* centre of expertise in that field.

6.3 Diversity

The element of ‘diversity’ was added to the SEP in July 2016, which was somewhat late in view of the (almost) finished self-assessments. The SEP mentions gender, age, ethnicity, organisational culture, and the diversity targets that research units have set themselves, and how these are reflected in their selection and assessment policy. These are immense tasks, which cannot and need not be carried out by everyone in the same way. But the message is clear: do not look away, and tackle it.

As diversity had not yet been included as an item of assessment in the SEP nor in the Discipline Protocol, it is not surprising that, apart from some passages, it gets little attention in the nine reports. However, the subject was regularly brought up during the discussions. Sometimes this concerned the flow of female talent, with the classic comment (regrettably) that despite the high numbers of female students, PhD candidates and lecturers/senior lecturers, the number of female professors is still lagging behind proportionally. Often the same comments also applied to researchers with a non-Dutch background and/or originating from other research cultures, including research of global subjects and, more broadly, efforts to make the internationalisation of legal research more visible. Occasionally, this also applied to the need for diversity as part of groups’ vitality. The SEP formulates it as follows: “It is precisely the presence of mutual differences that can act as a powerful incentive for creativity and talent development in a diverse research unit” (p. 9). Accordingly, the SEP argues that diversity is not a target, but a means to combine different perspectives and opinions. I couldn’t agree more. Academic literature on innovation shows that real breakthroughs and new insights are often achieved by bringing together people with quite different backgrounds (personalities, experience, age, disciplines). In my view, the focus when organising research programmes and research groups, as well as actual research, especially in relation to wide-ranging subjects, should be less on a collection of loose insights geared to individuals (the ‘zoo model’, with separate enclosures for separate species), and more on the ‘bio-diversity’ model (species diversity within one and the same ecosystem). Together these represent the equivalents of *multi-disciplinary* versus *interdisciplinary*.

Meanwhile, discussing the subject of ‘diversity’, with all its variations and dimensions, has become commonplace at all faculties; fortunately, it did not take the SEP to achieve this. As the protocols did not yet demand attention for diversity at the start of the assessment cycle, the faculties had no particular reason to supply information. This is evident in the nine reports as, although the term ‘diversity’ frequently appears in various meanings, no report contains a separate section or subsection on the subject. For the time being I can state that the subject is attracting interest, both for the present and for the future, with the key concepts in my view being the position of women, the mix of cultures and research cultures (‘inclusivity’) and the best way to achieve innovation. For the future, I would recommend recording the state of affairs in 2017 so that this can serve as a benchmark in the next assessment.

6.4 Integrity

Unlike diversity, the subject of ‘integrity’ did have to be assessed formally. As the reports and the discussions show, all the faculties have committed themselves to the relevant codes of conduct, in particular to the VSNU’s Code of Conduct for Academic Practice (*Gedragcode Wetenschapsbeoefening*) of 2004, which was revised in 2014 in cooperation with the KNAW. All faculties or their universities also have ethical committees, which may be contacted, either in advance in the event of research

questions with ethical implications, or afterwards in the event of integrity issues in or relating to the research itself. These issues are also dealt with in the SEP and the Discipline Protocol. Among the subjects mentioned are the handling of raw research data and data management, processing of research data, having protocols for data storage and paying attention to reproducibility. This latter issue has traditionally played a role in legal sub-disciplines such as legal sociology and criminology, but is now increasingly also doing so in other legal sub-disciplines because of a strong increase in both empirical research and research that is partly based on data collections. Some faculties have already adopted a policy document on data management and annexes. Other integrity issues relate to matters such as the dependence on specific funders of contract research, plagiarism, and how to deal with co-authorships and correct citations, with the keywords being ‘sensitization’ and the structural promotion and maintenance of that same sensitivity. Some faculties have already organised relevant workshops and seminars focusing on very concrete dilemmas and practical problems, and including issues such as a transparent research culture and the possibility of recourse to a confidential advisor or committee in the event of doubt. Whether all this is sufficient, only practice will tell.

6.5 Awarding ratings

The Discipline Protocol states that “legal research (...), unlike many other disciplines, does not rely on numerical data, but on judgements of substantive quality by assessment committees” (p. 13). This sentence is problematic for two reasons. First there is the assumption as regards the “many other disciplines”. In my view this is not correct, even though the word “many” provides some scope for interpretation. Discussions are being held in other disciplines, too, and it has become more widely accepted that reserving the term ‘quality’ for articles with a high impact factor and many citations does not constitute information that makes weighing superfluous, and moreover that such publications and citations only partially reflect the significance of research. This certainly applies to the other humanities and social sciences, while they, too, fall under the scope of the SEP, with its heightened attention for relevance to society.

Secondly: as regards the “numerical data” and the implicit assumption that the presence of sufficient numerical information makes weighing superfluous (“To measure is to know”), I have already argued how this applies to legal research. In various assessment discussions the question arose as to whether awarding ratings was in fact obligatory, alongside the underlying *narratives*, also given (section 1.2.4) that the present scale is quite rough and does not allow for intermediate categories. According to the SEP, however, ratings (or their equivalent) must be given for the parts for which a formal assessment is explicitly requested. After an internal consultation by e-mail between the nine chairpersons and the undersigned, it was concluded that the committees were not at liberty to omit ratings, but that they would communicate their unease and/or dissatisfaction in their reports, if desired.

This discussion is not only taking place in legal research but, according to information from one of the committee secretaries, also in other disciplines, such as in the recent assessment of mathematical research at Dutch universities. According to the chairperson of the relevant assessment committee – Professor Michiel Dekking from TU Delft, with whom I corresponded on this topic – the scale available has actually been reduced to two options, given “the fact that almost all Dutch mathematical research receives international recognition” and that, therefore, “a ‘3’ would not do justice to the performances”. He also pointed out that, in the present system, a whole institute can be recognised as world leading, while possibly only a few individuals in the case of a research unit of 10 FTEs and approx. 25 researchers may belong to the real global top. According to Dekking, the quantitative rating cannot, therefore, possibly do justice to the differences between universities and programmes, while he, too, argues that the *narratives* “present substantially more information than the quantitative scores”. It is interesting to hear this from a very different angle, i.e. from a discipline from which many a lawyer would expect to hear otherwise.

6.6 Benefits of the self-confidence of the discipline

Compared to the “many other disciplines” referred to in section 6.5, there is no need for legal research to be overly modest, neither in respect of its content, nor in respect of the state of the quality indicators. Much has been achieved that one can build on. This pride in the discipline’s uniqueness also touches upon something that has gained strength via the SEP and in current discussions on the role of universities, and has maybe even enjoyed a revival in the form of the growing attention for research’s relevance to society. In 1996, the Van Gerven Committee recommended that assessments “should henceforth be restricted to what is considered to be actual research work” (p. 10). Although the Committee gave valid reasons for this, driven by pragmatism and the need to distil a jumble of material in order to get to the core of practising legal research, its recommendation has been shown, after all, to be a product of its time.

Finally, there will always be a need for external supervision of performance, if ‘only’ from the taxpayer’s view, and even though assessment systems may differ. Research groups and faculty boards will also always be expected to be able to learn and adapt, given that the current challenges with regard to subject matter content and administration and management are not the same as those of twenty years ago and will most likely also be entirely different some decades from now. However, criteria such as ‘quality’, ‘relevance’ and ‘vitality’ will remain measuring points for all relevant output, even though the words may change again and new accents may be added. For the time being, I believe that, as far as assessments and indicators are concerned, legal research can confidently opt for ‘aligning where possible’ and for ‘uniqueness and self-confidence where appropriate’. Legal research has proven to be a long-term undertaking, the SEP marking a step in the direction of the discipline rather than the other way around. And this does not imply returning to the situation of some decades ago, but rather represents an embracing of progress and discipline specificity.

ANNEX

Author's CV

(relatively extensive so as to 'explain and substantiate' several arguments in this report).

Willem van Genugten (1950) is an Emeritus Professor of International Law at Tilburg University. His *alma mater* is Radboud University in Nijmegen, where he studied both law and philosophy (graduation 'with distinction' and *cum laude*) and defended his PhD thesis.

His previous positions (insofar as relevant here) have included being a visiting professor at the Minnesota School of Law, USA (2000-2014), co-founder and chair of the Center for Transboundary Legal Development at Tilburg University (2000-2001 and 2005-2013), Dean of the Tilburg Law Faculty (2002-2004 and 2010-2011), board chairman of the National Research School Human Rights (1998-2008), board chairman of NWO-WOTRO Science for Global Development (2006-2012), Dean of the Hague Institute for Global Justice (2011-2012) and the editor/editor-in-chief of the *Netherlands Yearbook of International Law* (2005-2016). He has also held several policy advice positions for the Dutch government and has been a member or chair of an NVAO audit team or verification committee on eight occasions.

At present, he is Professor of International Law by special appointment at North-West University, South Africa, and chair of the Dutch government's Knowledge Platform for Security and the Rule of Law, of the International Law Association's Committee on the Implementation of the Rights of Indigenous Peoples and of the Royal Netherlands Society of International Law. In 2016-2017 he headed departments of the schools of social sciences and economics of Tilburg University.

He has published on a series of international law subjects, both in the classical legal sense and from a multi/interdisciplinary perspective, with a special focus on human rights (in the widest sense of the word) and economic actors and factors.

In his Tilburg period (1985-2015), he was involved as an applicant/co-applicant in nineteen successful larger research applications and assignments for contract research (each involving over €50,000). So far, he supervised 32 PhD candidates on their way to the auditorium and has been on 55 other PhD committees at ten different universities (and honestly has no idea whether this number is comparatively high or low). In 2012, he was awarded an honorary doctorate by North-West University, South Africa for his "international legal stature", with reference being made, for example, to a then recently published leading article in the *American Journal of International Law* and to his "passion for promoting international scholarly relations".