Evaluation of research in context: an approach and two cases

Stefan P L de Jong, Pleun van Arensbergen, Floortje Daemen, Barend van der Meulen and Peter van den Besselaar

Science is increasingly heterogeneous, posing new questions for research evaluation. How can we evaluate the between scientific and societal quality of research, taking into account differences between research fields and between research groups? In this paper we present the findings of two case studies in fields where societal and scholarly output of research are highly intertwined (architecture and law). We analyze the nature of the two fields in terms of research areas and specific aspects of knowledge dynamics. This results in an approach and indicators for contextual research evaluation.

FOR A LONG TIME, societal and economic relevance of scientific research was taken for granted. In recent decades, however, changes in the societal role and position of science have ensured a more direct demand for relevant knowledge, which has been theorized in concepts such as *mode-2 knowledge production*, or the *triple helix* (Gibbons *et al*, 1994; Etzkowitz and Leydesdorff, 1998; Hessels and Van Lente, 2009). The *audit society* (Power, 1997) and *new public management* (Lane, 2000; Schubert, 2009) do not take societal relevance of research for granted; it has to be shown in grant applications and in evaluation of research programs and institutes. However, measuring the scientific and societal impact of research requires appropriate concepts and indicators.

First, with respect to the *scientific* impact, the role of publications and citations in high-impact journals has become dominant, especially in the natural and life sciences. This resulted in an abundance of citation-based indicators (Moed, 2005; Bornmann *et al*, 2009; Durieux and Gevenois, 2010), which are heavily debated even in these fields (Opthof and Leydesdorff, 2010; Van Raan *et al*, 2010; Bornmann and Mutz, 2011; Leydesdorff and Bornmann, in press; Waltman *et al*, 2011).

Second, research evaluation has a methodological bias towards the natural sciences (*Nature*, 2010), but the same methods are increasingly applied to technical sciences, social sciences, humanities and creative arts (Donovan, 2007; Martin *et al*, 2010). However, research output here is much more varied than scientific journal articles only, and also consists of books, proceedings, computer programs, designs and prototypes, etc. In these fields, the usefulness of citation-based indicators is even more questioned (Butler, 2007; Franceschet, 2010; KNAW, 2005; Nederhof, 2006; Nederhof *et al*, 2010; Van Leeuwen, 2006; TU Delft, 2009).

Third, the emphasis on *societal impact* adds a new dimension to research evaluation (Goransson *et al*, 2009; Gregersej *et al*, 2009; Krucken *et al*., 2009; Magnus Pålsson *et al*, 2009; Ca, 2009), which is still poorly addressed in evaluation practice (Nightingale and Scott, 2007). Several indicators for societal relevance have been proposed recently (Danish Council, 2006; Donovan, 2008; Grant *et al*, 2009), generally focusing on economic impact (HEGG *et al*, 2008) or health impact (Bensing *et al*, 2004).

Stefan P L de Jong, Pleun van Arensbergen, Floortje Daemen and Barend van der Meulen are all at the Science System Assessment Department, Rathenau Instituut, PO Box 95366, 2509 CJ, Den Haag, The Netherlands; Email: stefan, pleun, floortje, barend@rathenau.nl; Website: <www.rathenau.nl>. Peter van den Besselaar (corresponding author) is a professor at the Department of Organization Science and the Network Institute, VU University, Buitenveldertselaan 3, 1081 HV Amsterdam, The Netherlands; Email: p.a.a.vanden.besselaar@vu.nl; Website: <www.vandenbesselaar.net>.

Fourth, evaluating societal quality suffers from methodological problems, as it is difficult to *attribute impact* to specific inputs: The relation between knowledge and impact is complex and innovations are based on a variety of (knowledge) sources. Furthermore, it may *take years* before knowledge is applied and has impact. The mechanisms generating societal impact have hardly been studied (De Jong *et al*, forthcoming), therefore we do not understand how societal relevance and impact are generated and should be measured.

Nevertheless, some progress is visible. Based on the laboratory activity profiles approach (Callon *et al*, 1992; Larédo and Mustar, 2000), *positioning indicators* were proposed to relate performance of research groups[1] to their mission (Lepori, 2006; Lepori *et al*, 2008; Merkx and Van den Besselaar, 2008). Research groups have an (explicit or implicit) mission, specifying what kind of research and research outcomes are aimed at, and for which audiences. This can be narrow, for example, frontier research with top publications for peers and PhDs in the field as the only two output categories. It can also be broad when a research group also aims at contributing to innovation, professional work, policy development and public debate. Output will be much more varied and may include publications for professionals, policy reports, patents, and newspaper articles. In other words, research groups can have different (combinations of) audiences: the scholarly community, professionals, policy, companies, and the general public. For these audiences different types of research output are produced and should be taken into account in research evaluation. The evaluation of the *quality* of this heterogeneous output should be done against criteria defined by the respective audiences. In the scholarly domain, peers define quality and relevance of research output, and citations may be one indicator. In a similar way, the various societal audiences should assess quality and we need indicators for this too.

Consequently, *evaluating research in context* focuses on the interactions, that is communication and collaboration, between researchers and their scholarly and societal audiences (Spaapen *et al*, 2007; ERiC, 2010).[2] Research quality refers to all dimensions of this interaction and not only to 'impact' which may not yet be visible. Productive interactions also take place in agenda-setting, in collaborative research, in communication and disseminating of research outcomes, and in the use of knowledge. Productive interactions between researchers and the various audiences can be seen as 'proxy' for (future) impact.[3]

In this paper we test an approach for evaluating research in context, using the framework of productive interactions. In two case studies, we analyse agenda-setting, research collaboration, knowledge dissemination, and knowledge use (impact) which are *field-specific* and *group-specific*. This leads to the identification of the intended audiences and types of output, and to a set of indicators for scholarly and societal output and quality. The four steps are visualized in Figure 1.

Research evaluations should not be based on indicators that *a priori* cover the entire science system, as this has become too heterogeneous to be served by a *'one size fits all'* approach. In this paper we will show how a contextualized approach can be translated into indicators. More specifically, we will answer the following questions:

- What are the main characteristics of the research field under evaluation?
- What is the local context in which academic research groups are embedded, and how does this influence knowledge dynamics?
- What are the intended audiences and related types of output?
- What indicators for scholarly and societal output and impact can be derived from this?
- Finally, are the resulting evaluation approach and indicators useful?

1. Data and methods

Two fields (architecture within engineering and law within the social sciences) were selected as they combine several of the problems discussed above:
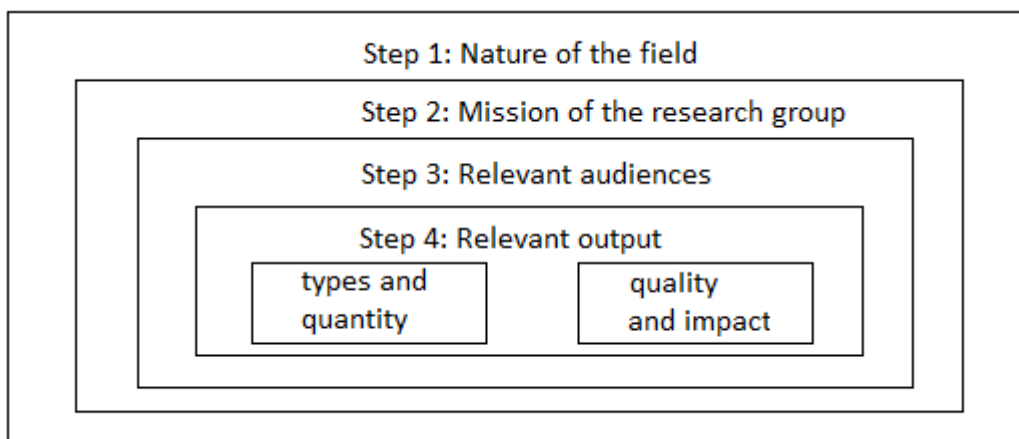

Figure 1. Approach to evaluation of research in context

1. Large heterogeneity;
2. Long-lasting and unsettled difficulties in determining indicators for scholarly output and quality (Van der Voordt, 1999; De Jong and Van der Voordt, 2000; Herweijer, 2003; Stolker, 2003; VSNU, 2005, 2007; Buruma, 2007; Franken 2008; QANU 2002, 2007);
3. Poor coverage in the Web of Science;
4. Strong orientation on societal relevance; and
5. A large overlap in scholarly output and societal output.

Moreover, both fields were to be evaluated soon, and our studies could be instrumental in the preparation of these evaluations. We aimed to include all subfields, as this would enable us to develop an approach that could be applied across the field. Based on these criteria, we selected one faculty of architecture (out of two) and seven faculties of law (out of nine). As the selected faculty of architecture is by far the largest, we have for both fields the majority of research included in this study.

Data were obtained from annual reports, websites, research proposals, self-assessment reports, evaluation reports, and studies reflecting on the specific knowledge dynamics in these fields. Data sources per case are summarized in Table 1.

Based on the findings of the document analysis, we conducted a first round of expert interviews with researchers and stakeholders to gain more insight into the nature and knowledge dynamics of the two fields. After the interviews, workshops with prominent researchers, faculty management and stakeholders were organized for both fields to present and test our findings about the nature of the field and the various types of audiences. In the workshops we started the discussion about output and quality indicators, related to the typology of audiences. The results were summarized and reported back to the participants for validation. In a second round of expert interviews with research leaders the indicators were refined, and then tested

Stakeholders were interviewed to gain more in-depth knowledge about the use and impact of different types of research output, the collaboration between researchers, and to gather input for a further test of the indicators

again through feedback and comments from the specialists.

Finally, stakeholders of the research groups were interviewed to gain more in-depth knowledge about the use and impact of different types of research output, the collaboration between researchers, and to gather input for a further test of the indicators. Throughout the process, we regularly had informal meetings with research directors of the faculties to receive feedback.

2. The architecture case

2.1 Research field context

The first element of the research context is the *heterogeneity* of the field. The faculty of architecture distinguishes four main subject areas, related to different disciplines: the design of buildings related to art; social study of urban and regional processes and structures; building technologies; economics and management of building processes and the existing stock of real estate. These areas are rather heterogeneous themselves, as they deploy different research approaches. First, *evaluation research*, the empirical study of how buildings, cities and regions function. Second, research about the *historical* development of design ideas and practices. Third, *conceptual* (exploratory and experimental) research aiming at innovative and revolutionary concepts, manifestos, visions and materials for the architecture, urban planning and building. Finally *design research* for professional practice, collecting knowledge needed to find optimal solutions for a specific building assignment (Table 2).

Second, much stronger than other technical sciences, artistic, cultural and *normative perspectives* play an important role in architecture, especially within conceptual research. These specific characteristics of architecture and building research are covered under the label 'research by design': research during and by designing, covering the four approaches distinguished above. It is often discussed whether 'research by design' meets a level of scientific rigor comparable to the natural sciences. In

Table 1. Data

	Architecture (1 faculty out of 2)	Law (7 faculties out of 9)
Annual reports	√	√
External evaluations	√	√
Self-evaluations	√	√
Research programme proposals	√	√
Websites of faculty and research groups	√	√
Interviews with tenured staff	15	23
Interviews with stakeholders	14	20
Workshops with researchers and stakeholders	2	2

Table 2. Research in architecture, two typologies

Typology 1. Research area-based			
Architecture research	**Urbanism**	**Building technology**	**Real estate and housing**
Related to arts and humanities	Related to sociological research	Related to the natural and technical sciences	Related to management studies
Study of buildings, the built environment and theory of architectural design	Study of existing and possible (designed) spatial objects, like patterns of spatial development, urban areas, cities or buildings	Study of technical and material dimension of buildings	Study of the building process (initiation, preparation, development, design and construction), its dynamics and the management of real estate
The creative component is large			

Typology 2. Research approach-based			
Evaluation research	**Historical research**	**Conceptual research**	**Practical design research**
Empirical study of functioning of objects and processes	The understanding and explanation of the development of designs and design process while paying attention to site characteristics	Exploratory and experimental research, aiming at innovative, revolutionary concepts, manifestos and visions about the built environment	Research for professional practice
Study of societal effects (and quality) of buildings and spatial structures			Produces knowledge needed to find optimal solutions for a certain building assignment

architecture, however, the concept of scientific rigor itself has to be balanced with specific contextual demands of a normative nature, such as reflection and creativity in the design process, and the normative appreciation of problems in the built environment. Consequently, a tension exists between traditional criteria used by funding agencies and in research evaluation, which seem not to grasp the necessary normativity of architecture research.

Third, architecture research has a strong *relation with practice*, since practices and practitioners are not only an important audience, but also the object of study. The object of research is man-made and in many instances even research-made. For example, the study of problems within an urban environment may result in theories influencing architectural and urban planning practices — which in turn can be studied. The object of research, and the way it is approached, change as a result of knowledge development. We can recognize the intertwining of research and practice in all the four types of research: studies derive objects and processes from practice and aim at contributing to improving practice.

Intertwining with practice is also reflected by a large number of full professors who are very actively involved in architectural practice. Top architects are offered part-time positions as professors. By doing so, knowledge flows from practice to science and vice versa through a single person. The faculty currently employs 26 part-time professors[4] who also hold positions at well-known national and international architectural offices (Avermaete *et al*, 2010).

The heterogeneous, normative and practice-orientated nature of architectural research points at a variety of audiences of architectural research. Briefly, four types of audiences can be distinguished in this case: peers, professionals (architects, urban planners), companies (e.g. contractors and housing corporations) and government agencies (Table 3).

Table 3. Stakeholders in architecture research

Evaluation research	Historical research	Conceptual research	Practical research
Peers (researchers)	Peers (researchers)	Peers (researchers)	Client who commissioned the work
Professionals: • Architects • Urban designers		Professionals: • Architects • Urban designers	
Companies: • Spatial planning firms • Contractors • Real-estate developers		Companies: • Spatial planning firms • Contractors • Real-estate developers	
Governments: • Municipalities • Provinces • National government		Governments: • Municipalities • Provinces • National government	

2.2 Aspects of knowledge dynamics

What do these characteristics mean for the four dimensions of knowledge dynamics: agenda-setting, research collaboration, knowledge dissemination, and impact?

Agenda-setting A tension exists between orientation on practice and academic freedom. Researchers emphasize that societal stakeholders (such as companies or government) may have specific interests, and therefore should not determine research agendas. Furthermore, problems from practice cannot always be translated into scholarly research and therefore should not always automatically be incorporated in research agendas. In other words, research and practice should interact but this does not imply that researchers are kept on a short leash by stakeholders. Nevertheless, researchers do interact with stakeholders when establishing research agendas.

So, given this tension, how is the research agenda determined? On the one hand, important issues and questions arising in practice (the object of research) influence the research agenda. Some researchers are inspired by societal concerns and take into account governmental policy and societal trends (such as sustainability) when formulating a research program. Others more directly involve their stakeholders when formulating the research agenda. This is often done through recruiting part-time professors who remain (or were in the past) actively involved in architectural and building practice. In other cases it is done through workshops, symposia or annual meetings with stakeholders to discuss the research agenda and keep in touch with what is considered topical and important by practice.

It should be emphasized that researchers also seek to influence stakeholders' policy agenda . Many interviewees explicitly mentioned their role in shaping agendas of societal actors. The research groups consider it to be an important societal responsibility to make societal questions explicit, bring actors from society together to formulate agendas, and critically reflect on agendas of societal actors.

Research collaboration We have identified three types of collaboration with stakeholders in architecture research. They differ in their degree of involvement and consequently in the degree to which knowledge is exchanged. In *commissioned research*, a research group aims at solving a problem posed by for instance a government agency or a housing corporation. The research question generally is jointly elaborated but during the research process, collaboration is limited. In the majority of cases, it is an exchange of knowledge for money. A second form of collaboration is the *long-term funding* of targeted research. A societal actor may acknowledge the long-term importance of a research theme and support it financially. Although the actor is interested in the content of the program, it is not involved in the research itself. Influence is exerted from a distance. The third most extensive form of collaboration is *joint research*. This type of collaboration is characterized by a two-way stream of knowledge between researchers and societal partners. The most common form is attracting professors from practice. Another form is a long-term partnership between a research group and a societal organization or company.

Knowledge dissemination Dissemination of architecture research uses a large variety of output types for different audiences (see Table 4 for a summary). Our study indicates that texts are the main form of output. Publications in peer-reviewed journals as well as in professional magazines are considered important, as are policy reports and books. Most research groups prefer professional journals or books, because these types of output are more suited to reach their societal and scholarly audiences. This is confirmed by an analysis of architectural publications in WoS-indexed journals. Nine well-known departments[5] of architecture together have had 462 publications (78% articles) between 1987 and 2009, which is only six papers per year — although there has been a rise in recent years. Clearly, this kind of research output is only marginally important in the field of architecture.

Our study also showed the role of non-textual output, such as software tools, drawings, computer animations, scale models, and prototypes of buildings

Table 4. Classes of output by types of architecture research

Evaluation research	Historical research	Conceptual research	Practical research
ISI publications	Books	Conference proceedings	Exhibitions (and catalogues)
Professional publications	Exhibitions (and catalogues)	Exhibitions (and catalogues)	Designs: • Drawings • Computer animations • Scale models
Policy reports		Designs: • Drawings • Computer animations • Scale models	
Tools			Prototypes: • Buildings • Constructions
		Prototypes: • Buildings • Constructions	

as well as constructions. Visualizations in turn can be discussed in exhibitions, lectures, debates, colloquia, seminars and conferences. Some of these events do result in publications such as exhibition catalogues or conference proceedings.

Last but not least is the dissemination of research results by people. As mentioned before, part-time professors serve as an important link between practice and science. They not only suggest relevant research subjects, but also communicate research results to practice. Furthermore, researchers' advisory and consultancy activities serve to disseminate knowledge to governments, companies, non-governmental organizations and to society at large.

As a consequence of the strong orientation on practice, communication with local stakeholders highly matters to research groups. Researchers therefore prefer national (Dutch language) professional journals to communicate with stakeholders, as stakeholders are not expected to read international scientific journals. Additionally, researchers feel that papers about local problems can hardly be published in international journals.

Impact   Based on the dominant role of societal audiences, one would expect to easily find examples of research output with impact in society. Interviewed researchers, however, could hardly give an overview of the impact of their research. And the direct (first-order) stakeholders do not always provide feedback about the use of research results. In other cases, impact is generated through indirect (second- or third-order) stakeholders, which makes keeping track of impact even more difficult.

Nevertheless, stakeholders proved capable of informing us about the different forms of impact of architectural research. For instance, a housing corporation introduced a new management model that came out of a research project, changing its everyday managerial practice. Results of another research project were referred to in a letter to parliament by the Minister of Housing. A third example of impact is the use of new building typologies by a Dutch municipality for city expansion, which

In the interviews and workshops we tried to find out what the various intended audiences of the research output consider as quality

changed the way urban areas and buildings are designed and constructed.

2.3 Quality indicators

In Table 4, we listed types of research output that can be measured. What indicators for the quality of the output might be useful? As argued in the introduction to the paper, the audiences (stakeholders) play a crucial role in defining quality. In the interviews and workshops we therefore tried to find out what the various *intended audiences* of the research output consider as quality. And, quality indicators should not be restricted to *impact*, but should cover all phases of the knowledge production process. Table 5 summarizes the quality indicators that we found in this case. Some of them can be easily quantified, for others this is much more difficult and less appropriate. However, as we focus on *positioning indicators* that compare quantity and quality of the research with the mission of the evaluated research group, this is not a problem. The aim of our approach is *research evaluation in context*, and not a ranking of research groups or programs. As Section 3 will demonstrate, similar quality indicators could be formulated for the law case.

3. The law case

3.1 Research field context

As in architecture, law is a *heterogeneous* research field, which results in a variety of field classifications.

Table 5. Indicators for evaluating societal quality of architecture research

Agenda-setting	Collaboration	Dissemination	Impact
1. Societal issues explicitly addressed in research	1. Commissioned research by societal actors	1. Scholarly and professional publications, including in local language	1. Convincing examples of use of outcomes of research
2. Occasional/structural interaction with stakeholders to establish relevance	2. Earmarked/structural funding related to societal theme	2. Technologies, artefacts, exhibitions, standards, designs	2. Satisfaction/recognition of alumni and stakeholders
3. Relevant experience of researchers as practitioner in a societal domain	3. Actual collaboration in research, testing and evaluation with stakeholders	3. Advisory/ consultancy roles	3. Substantial returns or economic value of outputs of research
4. Positive evaluations or external funding related to societal/commercial issues		4. Popularization, contribution to societal debate	4. Visibility in the public debate/media rankings
		5. Education, training of professionals, graduates	

Textbooks distinguish four areas of law: private law, constitutional law, administrative law and criminal law (Janssen *et al*, 1999; Cliteur, 2000). A closer look at the research programs of law faculties, however, shows that law research does not neatly fit into this classification. The organization of departments and research groups follows different logics. Moreover, the four areas of law do not cover the so-called 'meta-juridical' studies in which law is being studied from the perspective of other fields, like sociology or philosophy. For classifying research programs, it is more useful to distinguish between private law; constitutional and administrative law; criminal law and criminology; international and European law; and meta-juridical studies.

Independently from this, two major types of research should be distinguished: doctrinal research and empirical research. Doctrinal (humanities-oriented) research consists of descriptions and analyses of legal sources, and aims at uncovering the internal structure of law, such as the underlying (philosophical) assumptions, its internal coherence and the lack of it, and the way it is interpreted in jurisdiction. Empirical research, on the other hand, focuses on the way law and legal institutions function in society. This type of (social science) research is based on systematic empirical observations (Table 6).

Second, doctrinal law research has a strong normative character. The researcher's opinions often resonate in the research, which aims at normative judgments. In jurisprudence the primary question is: 'Should it be like this?' instead of: 'Is it really like this?' (Stolker, 2003; Kwakman, 2005). Consequently, many scholars tend to be more concerned with 'ought' than with 'is', more with improving law than with explanation. Improvement is pursued mainly by ordering, comparing and interpreting current law and legislation.

Third, the object of legal research is law and the legal system. The object is 'man-made' and subject to change caused by human action: 'Law is what people agreed on to be law.' It is not a static or secluded research object; law is studied and practised simultaneously. This reflects a main characteristic of law research, which some people consider as strength, others as pitfall: the strong intertwining with legal practice (e.g. Stolker, 2003). The legal system and its legal norms and rules constitute the primary research object. These are ordered, described, compared, interpreted and commented upon, in order to ensure better jurisdiction. According to many researchers, the most important task of academic law research is to inform (and improve) practice. This strong orientation towards practice is reflected in the large number of *professional* publications compared to *scientific* publications. The first mainly aim at explaining and clarifying for legal practice, while the latter particularly focus on communicating new insights. However, a lot of professionals also use scientific publications.

Another indication of the strong intertwining of research and practice is that a large share of the researchers in law schools, particularly professors, occupies positions in the legal practice, next to their position at the university, which is considered necessary within the field of law research. In order to conduct relevant good-quality research, researchers need to maintain *feeling* with practice. For example, besides being a professor they hold a position as a lawyer, deputy judge or legal counsellor. This is similar to architecture research, but in sharp contrast

Table 6. Research in law, two typologies

Typology 1. Research area-based

Private law	Constitutional and administrative law	Criminal law and criminology	International and European law	Meta-juridical studies
Studies legal relations between individuals (e.g. contract law, property law, family law, commercial law, and inheritance law)	Studies relationships between the state and individuals, and between different branches of the state Studies agencies' roles and power	Studies criminal behaviour, its causes, prevention and sanctioning from a legal perspective (criminal law) and from a psychological or sociological perspective (criminology)	Studies the European and international aspects of law, including the legal relations between countries Also comparative study of national legal systems	Studies the legal system from a sociological, economic, and philosophical perspective

Typology 2. Research approach-based

Doctrinal research	Empirical research
Descriptions and analyses of legal sources (e.g. legislation, jurisprudence) from the perspective of the legal system	Empirical studies of the functioning of the legal system
The aim is to structure law, indicate inconsistencies, and to add, adjust and improve where needed	The focus is on societal dynamics, functioning and effects of law and legal institutions

to many other fields, where this is regarded as undesirable. Because of this strong interdependence, it is often difficult and even considered irrelevant to distinguish between scientific and professional research output and stakeholders. It is hard to disentangle the different domains and evaluate the scholarly part of research output independently from it societal quality.

Fourth, and also similar to architecture, legal research has a rather strong *national orientation* — a worldwide phenomenon. For legal scholars in the Netherlands, *Dutch* jurisprudence and legislation constitute the key context in which they conduct research. International and comparative law research constitutes only a small part of the discipline. The Dutch legal professionals make up the main audience, resulting in Dutch as the main language of publication. Although the importance of publishing in English is generally acknowledged, it is claimed that an important part of Dutch legal research can simply not be published in English, as the Dutch language has its own specific legal concepts which would lose their meaning when translated (VSNU, 2007). Furthermore, international publications would require much more contextual information (about the Dutch legal system) to make a publication understandable.

Finally, despite the typical Dutch inclination towards programming of research, the study of law is, and has traditionally been, a field in which mainly *individual and small-scale* research is conducted. As no expensive instruments or laboratories are required for this type of research, researchers are not forced to cooperate — mutual dependency is low (Whitley, 2000). In contrast to architecture, law research is highly disciplinary, reinforcing its individualistic character. Although multidisciplinary research is increasingly considered to be valuable, it only evolves slowly. Legal scholars mainly stay within their own (sub)field and therein they all have their own specialism.

This analysis leads to a distinction between the domains or audiences in which law research is relevant (Table 7). Within the field, three domains are distinguished: the (inter)national scientific domain (peers); the domain of the public and private legal practice (professionals); and the political and societal domain (policy-makers and the general public).

3.2 Aspects of knowledge dynamics law research

Agenda-setting Academic researchers determine the research agenda. However, because many researchers are also involved in legal practice, the relation between research and practice is strong. Researchers are perfectly aware of what is taking place in practice, where the knowledge gaps are, and which important issues need to be investigated. Research questions often involve problems professionals encounter in their daily work, for instance when they are asked to provide legal advice in new or uncommon situations, where existing legal rules cannot directly be applied. Current societal problems also play an important role in research programs. Mission statements of the different law research programs describe the importance of studying *societal problems* and they reflect the high value ascribed to societal relevance.

Furthermore, to some extent research questions are influenced by external parties (from the political and business world) through commissioned research. According to our interviewees, external parties' impact on the research agenda remains small, as researchers generally have the freedom to elaborate research questions to make them better relate to existing research programs.

Research collaboration Although the actual conducting of research lies primarily in the hands of researchers, important stakeholders are involved in law research. For example, stakeholders may provide input for setting up the research project, and for the formulation of research questions, for example whenever the Ministry of Justice requires an analysis of legislation. The most important form of research-practice collaboration is researchers often conducting research 'in the field'. Many researchers not only study law; they work with it at the same time. As in architecture, this means individuals carry knowledge flows from research to practice and vice versa.

A second form of collaboration is at the institutional level: research centres based on partnerships between universities and private parties such as law firms, notary offices, pension funds, legal departments of large enterprises and financial institutions. These private partners enable scholars to

Table 7. Stakeholders in law research

(Inter)national science	(Public and private) legal professionals	Politics and society
• Peers (also scientists from non-legal fields) • Students	• Investigation services (e.g. police) • Advocacy • Public prosecutor • Jurisdiction • Council of State • (Part of) public administration • Notaries • Insurers • Mediators	• Government • Ministries • Policy-makers • Governmental advisory bodies (e.g. Council of State) • European Committee • Non-government organizations • Non-government advisory bodies

Many researchers not only study law; they work with it at the same time. As in architecture, this means individuals carry knowledge flows from research to practice and vice versa

conduct research in the companies' practices. These companies also offer internships to students, which enables them to observe legal knowledge in practice. In return private partners have the opportunity to 'scout' future juridical talent and access relevant and useful scientific knowledge. To elaborate on this last point, there seems to be a need for more *scientification* of legal practice. For example, jurisdiction established several chairs at Dutch universities in order to fulfil the need for scientific deepening of jurisdiction.

Knowledge dissemination Legal research focuses on different audiences such as peers, professionals, politicians and the public (Table 7). Researchers do interact with these audiences in terms of a varied research output (Table 8).

The large variety of types, audiences and functions of publications requires a subdivision. The most important part consists of scientific and professional publications. We take them together as it is often difficult to differentiate between these two types. The majority of the (Dutch language) law journals are being read by both the scientific and professional community. Other categories of publications are monographs (highly valued within law research), dissertations (mainly in the form of a monograph, rarely as a compilation of articles), conference proceedings, preliminary advices, inaugural lectures, (advisory and policy) reports and popularizing articles.

The boundary between scholarly and professional publications is fuzzy and disputed. For example, the *annotation* is a concise commentary on a judicial verdict and is directed to both legal professionals and researchers. Although they are regarded as very valuable output by legal practitioners, annotations are not always considered scholarly output.

Another form of textual output are *tools*, *manuals* and *codes of conduct*, which can be utilized in legal practice and elsewhere in society. Examples of these types of output are guidelines for the impartiality of judges and the code of conduct for the treatment of injury claims. Many of these tools are published on the internet, making them accessible to a wide audience.

Furthermore, contributions to national and international *conferences*, symposia, lectures and expert meetings are also considered to be important types of research output. The audiences may vary: scholars, policy-makers and members of parliament, professional lawyers, or the general public.

Another important way to disseminate research results is via the *mobility of people*. Part-time researchers can directly disseminate and implement academic knowledge into practice. Therefore, sidelines in legal practice are considered research output as well.

Dissemination can also be realized via *membership* of (inter)national scientific committees, networks and editorial staff of journals and membership of political and societal advisory and policy committees. The latter affiliation enables academic knowledge to flow directly into professional practice and society.

Finally, post-academic education is an important way of disseminating academic knowledge via people. Researchers provide postgraduate education for jurists in favour of their legal practice. By organizing interactive seminars, researchers are both disseminating academic knowledge and being informed by practice.

Impact The variety of research output described above can be considered as the instruments used by researchers and research groups to translate their mission into scholarly and societal impact. As in architecture, legal researchers too found it difficult to indicate the impact of their research and link signs of impact to specific research projects despite the fact that legal research and practice are strongly intertwined, and impact may be realized rather directly. Nevertheless, researchers are sometimes aware of the practical use and influence of their research. Research regularly leads to parliamentary questions, and to changes in rules and legislation. For instance the Council of State biweekly discusses the latest

Table 8. Classes of output categories in law research by audience

(Inter)national science	(Public and private) legal professionals	Politics and society
• Scientific and professional publications • Membership of scientific committees, networks and editorial staff • Contribution to conferences and symposia	• Scientific and professional publications, reports, manuals • External function in legal practice, advisory body • Membership of editorial staff • Contribution to conferences and symposia • Post-academic education	• Professional and popularizing publications • Membership of political and societal committees • Contribution to conferences and symposia

journal papers and annotations, and determines whether this coincides with the current jurisprudence or whether adjustments have to be made. However, use is also often unnoticed, as legal pleas and judgments lack references. Nonetheless, many researchers are actively involved in legal practice and in advisory committees, where they can implement knowledge and directly observe impact.

3.3 Quality

In a similar way as for architecture, output indicators were developed. Based on the interviews and workshops with researches and societal stakeholders, we determined what different audiences consider to be 'quality'. These quality criteria relate to the several phases of the knowledge production process, and not only to impact. It is relevant to emphasize that the bibliometric databases do not play a role in law, for instance because the WoS is heavily biased towards US law journals and US research, and therefore cannot be used in research evaluations of law research elsewhere (Moed, 2005). The indicators are listed in Table 9. Again, some indicators can be easily quantified, while other indicators are qualitative.

4. Evaluation research in context: conclusion and discussion

In this paper we introduced an approach for evaluating research groups in their disciplinary and local context. Frame of reference is the *mission* of the research group, as this defines what the group is expected to accomplish. First, the different scholarly and societal *audiences* (or stakeholders) of a research group are identified.

Second, in interaction with researchers and their audiences, the specific *types of output* for these audiences, and the *interactions* between researchers and audiences are identified.

Third, *indicators* for research output, quality, and impact were developed.

Fourth, in contrast to what is done usually, we do not restrict quality to visible impact. A more general quality concept was introduced that takes into account the quality of communication and collaboration between researchers and their audiences: productive interactions that may result in impact at some future moment.

Finally, stakeholders are needed to assess societal impacts, comparable to scientific peers who are able to evaluate scientific impact. Carefully selected local stakeholders can be valuable in determining societal impact. Stakeholders were able to indicate in what way research is relevant for them, how they productively communicate with researchers, and how they use the results in their daily work — even if this cannot always be measured through formal and explicit references to research output.

In order to develop the approach, we started with two practice-oriented research fields, architecture and law, as these have very heterogeneous audiences and research output. Our two case studies showed the practical *usability*. Although the information required was not always easily available, the cases illustrate that it can be collected.[6] And this does not lead to huge amounts of paperwork and excessive workload, as is sometimes suggested (Grant *et al*, 2009). The main result, however, was in the architecture case. There, the recent formal research evaluation report was based on the approach presented in this article (Avermaete *et al*, 2010). The same holds for recent proposals for evaluating engineering research (KNAW 2010).

Interactions with non-academic stakeholders are an important way of circulating knowledge between science and society. The intensity of the collaboration informs us about the type and amount of knowledge that is circulated. On top of that, collaborations are an indication of societal quality. In both fields, collaboration with stakeholders can be considered a significant way to circulate knowledge between the different domains of science, professional practice, politics and society. Although stakeholders do not play a substantial role in research practice

Table 9. Indicators for evaluating societal quality of law research

Agenda-setting	Collaboration	Dissemination	Impact
1. Societal issues explicitly addressed in research	1. Commissioned research by societal actors	1. Scholarly and professional publications, including local language	1. Convincing examples of use of outcomes of research
2. Occasional/structural interaction with stakeholders to establish relevance	2. Partnerships between universities and external parties/stakeholders	2. Books, monographs, pre-advices, annotations	2. 2nd editions of books
3. Relevant experience of researchers as practitioner in a societal domain	3. Academic researchers enabled to conduct research 'in the field' by working in practice concurrently	3. Guidelines, tools, manuals, codes of conduct	3. Pre-advices
		4. Contributions to national conferences/symposia/expert meetings	4. Commissioned research
		5. Advisory and consultancy roles	5. Visibility in the public debate/public media
		6. Training of professionals	

itself, they do enable researchers to conduct research 'in the field' and remain in close contact with practice.

In the two case studies, researchers are at the same time practitioners, and research is often individual and small-scale. These characteristics influence the nature of the researchers–stakeholder interactions, the dynamics of agenda-setting, collaboration, and use of knowledge. The concept of productive interactions needs to be studied more systematically. We are currently studying other fields, where research is large-scale and more integrated. This leads to different types of interaction, often more indirect and 'networked'. Consequently, other quality and impact indicators will be required for these fields.

We end with two general issues. First, evaluation is often aiming at the ranking of research groups. In contrast, the proposed approach is meant for the evaluation of the performance of a research group against its own mission. As missions differ, the result of evaluations may not be easy to compare. Only where missions and research fields are sufficiently similar is comparison meaningful. However, given the large number of goals, audiences, and types of output, every reduction of performance to a single figure would be meaningless. Researchers themselves play a significant role in the dissemination of their output. Since many of them are practitioners at the same time, new knowledge can immediately be applied in practice. Additionally, researchers are often members of scientific, professional, societal, and policy advisory committees and they provide post-academic education. This also creates the possibility of implementing research results directly into the different domains.

Second, our study has shown a large variety of stakeholders, types of collaboration, and forms of dissemination within both architecture and law, illustrating the complexity and heterogeneity of the contemporary science system. Classifications of research into two 'modes' (Gibbons *et al*, 1994), or into four 'quadrants' (Stokes, 1997), seems too general for analysing the dynamics of the science system and particularly for research evaluation. The frequently discussed change of 'the relations between science and society' has resulted in a large variety of types and contexts of scholarly research, and appropriate evaluation approaches are needed that reflect this heterogeneity. This paper is an empirical contribution. Current (and future) work focuses on refinement of the approach and on testing it in other contexts and disciplines.

Acknowledgments

Notes

1. Where we use 'research group', one may also read 'institute' or 'program'.
2. We focus on evaluation for organizational learning and not for (anyhow not useful for) ranking exercises.
3. Increasingly framed in terms of the role of user engagement in research collaboration and agenda setting (Donovan and Butler, 2007).
4. At the full–professor level.
5. MIT, ETH Zurich, University of Cambridge; Technical University Delft; Technical University Eindhoven; University of Sheffield; University of Reading; Ghent University.
6. Evaluation requires skills and guidelines alongside the evaluated researchers. To support learning, a guideline was developed to be used in the regular research evaluation (ERiC, 2010).

References

Avermaete, T, H Coolen, R Docter *et al* 2010. *Architecture and the Built Environment: Research in Context 2003-2009*. Delft: TU Delft Architecture.

Bensing, J M, W M C M Caris-Verhallen, J Dekker *et al* 2004. Doing the right thing and doing it right: toward a framework for assessing the policy relevance of health services research. *International Journal of Technology Assessment in Health Care*, **19**(4), 604–612.

Bornmann, L and R Mutz 2011. Further steps towards an ideal method of measuring citation performance: the avoidance of citation (ratio) averages in field-normalization. *Journal of Informetrics*, **5**(1), 228–230.

Bornmann, L, R Mutz and H D Daniel 2009. Do we need the H index and its variants besides standard bibliometric measures? *Journal of the American Society of Information Science and Technology*, **60**(6), 1286–1289.

Buruma, Y 2007. Rechtswetenschap en rechtspraktijk. *Nederlands Juristenblad*, **17**, 1043.

Butler, L 2007. Assessing university research: a plea for a balanced approach. *Science and Public Policy*, **34**(8), October, 565–574.

Butler, L and M S Visser 2006. Extending citation analysis to non-source items. *Scientometrics*, **66**(2), 327–343.

Ca, T N 2009. Reaching out to society: Vietnamese universities in transition. *Science and Public Policy*, **36**(2), March, 91–95

Callon, M, P Larédo, P Mustar *et al* 1992. Defining the strategic profile of research labs: the research compass card method. In *Science and Technology in a Policy Context*, ed. A F J van Raan. Leiden: DSWO Press.

Cliteur, P B 2000. *Inleiding in het recht*. Deventer; Kluwer.

Danish Council 2006. *A Tool for Assessing Research Quality and Relevance*. Copenhagen: Danish Council for Research Policy.

De Jong, S P L, B van der Meulen, J Spaapen and P van den Besselaar (forthcoming). *Productive Interaction and Social Impact: a Literature Review*.

De Jong, T M and D J M van der Voordt 2000. *Criteria voor wetenschappelijk ontwerp*. Available at <http://team.bk.tudelft.nl/Publications/2000/Ways%20to%20study%20prelininary%20Dutch%20versions/02Criteria%20HYBRIDE04.htm>, last accessed 28 October 2010.

Donovan, C 2007. The qualitative future of research evaluation. *Science and Public Policy*, **34**(8), October, 585–597.

Donovan, C 2008. The Australian Research Quality Framework: a live experiment in capturing the social, economic, environmental and cultural returns of publicly funded research. *New Directions for Evaluation*, **2008**(118), 47–60.

Donovan, C and L Butler 2007. Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': an economics pilot study. *Research Evaluation*, **16**(4), December, 231–242.

Durieux, V and P A Gevenois 2010. Bibliometric indicators: quality measurements of scientific publication. *Radiology*, **255**, 342–351.

ERiC 2010. *Evaluating the Societal Relevance of Academic Research: a Guide*. The Hague: ERiC.

Etzkowitz, H and L Leydesdorff 1998. The endless transition: a 'Triple Helix' of university-industry-government relations. *Minerva*, **36**, 203–208.

Faculty of Architecture, Delft University of Technology 2007. *Assessment of Research 2003-2006*. Delft, Faculty of Architecture, Delft University of Technology.

Franceschet, M 2010. A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, **83**(1), 243–258.

Franken, H 2008. *Rechtsgeleerdheid in de rij der wetenschappen*. Amsterdam: KNAW.

Gibbons, M, C Limoges, H Nowotny *et al* 1994. *The New Production of Knowledge*. London: Sage.

Göransson, B, R Maharajh and U Schmock 2009. New activities of universities in transfer and extension: multiple requirements and manifold solutions. *Science and Public Policy*, **36**(2), March, 157–164.

Grant, J, P B Brutscher, S Kirk *et al* 2009. *Capturing Research Impacts*. Cambridge: RAND Europe.

Gregersen, B, L T Linde and J G Rasmussen 2009. Linking between Danish universities and society. *Science and Public Policy*, **36**(2), March, 151–156.

HEGG *et al*, Health Economics Research Group, Office of Health Economics and Rand Europe 2008. *Medical Research: What's it Worth? Estimating the Economic Benefits from Medical Research in the UK*. London: HEGG.

Herweijer, M 2003. *Juridisch onderzoek*. Available at <http:/irs.ub.rug.nl/ppn/275103943>, last accessed 4 December 2008.

Hessels, L K and H van Lente 2009. Re-thinking new knowledge production: a literature review and a research agenda. *Research Policy*, **37**, 740–760.

Janssen, O, E van de Vliert and C Veenstra 1999. How task and person conflict shape the role of positive interdependence in management teams. *Journal of Management*, **25**(2), 117–142.

KNAW 2005. *Judging Research on its Merits: an Advisory Report by the Council for the Humanities and the Social Sciences Council*. Amsterdam: KNAW.

KNAW 2010. *Quality Assessment in the Design and Engineering Disciplines: a Systematic Framework. Advisory Report of the Council for Technical Sciences, Mathematics, Informatics, Physics, Astronomy and Chemistry*. Amsterdam: KNAW.

Krücken, G, F Meier and A Müller 2009. Linkages to the civil society as 'leisure time activities'? Experiences at a German university. *Science and Public Policy*, **36**(2), March, 139–144.

Kwakman, N J M 2005. *Het gelijk van Knigge*. Available at <http://rechten.eldoc.ub.rug.nl/FILES/root/Feestbundels/liberA micorumKnigge/hfdst21/hfdst21-kwakman.pdf>, last accessed 4 December 2008.

Lane, J E 2000. *New Public Management*. London: Routledge.

Larédo, P and P Mustar 2000. Laboratory activity profiles: an exploratory approach. *Scientometrics*, **47**(3), 515–539.

Lepori, B 2006. Methodologies for the analysis of research funding and expenditure: from input to positioning indicators. *Research Evaluation*, **15**(2), August, 133–143.

Lepori, B, R Barré and G Filliatreau 2008. New perspectives and challenges for the design and production of S&T indicators. *Research Evaluation*, **17**(1), March, 33–44.

Leydesdorff, L and L Bornmann 2011. How fractional counting of citations affects the impact factor: normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology*, **62**(2), 217–229.

Martin, B, P Tang, M Morgan *et al* 2010. *Towards a Bibliometric Database for the Social Sciences and Humanities: a European Scoping Project*. Brighton: SPRU.

Merkx, F and P van den Besselaar 2008. Positioning indicators for cross-disciplinary challenges: the Dutch coastal defence research case. *Research Evaluation*, **17**(1), March, 4–16.

Moed, H F 2005. *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Nature 2010. **465**, 845–974.

Nederhof, A J 2006. Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics*, **66**(1), 81–200.

Nederhof, A J, T N van Leeuwen and A F J van Raan 2010. Highly cited non-journal publications in political science, economics and psychology: a first exploration. *Scientometrics*, **83**(2), 363–374.

Nightingale, P and A Scott 2007. Peer review and the relevance gap: ten suggestions for policy-makers. *Science and Public Policy*, **34**(8), October, 543–553.

Opthof, T and L Leydesdorff 2010. Caveats for the journal and field normalizations in the CWTS ('Leiden') evaluations of research performance. *Journal of Informetrics*, **4**(3).

Pålsson, C M, B Göransson and C Brundenius 2009. Vitalizing the Swedish university system: implementation of the 'third mission'. *Science and Public Policy*, **36**(2), March, 145–150.

Power, M 1997. *The Audit Society: Rituals of Verification*. New York: Oxford University Press.

QANU 2002. *Onderzoeksbeoordeling Rechtsgeleerdheid*. Utrecht: VSNU.

QANU 2007. *Mid-term Research Assessment Architecture and Building Science*. Utrecht: VSNU.

Schubert, T 2009. Empirical observations on new public management to increase efficiency in public research – boon or bane? *Research Policy*, **38**, 1225–1234.

Spaapen, J, H Dijstelbloem and F Wamelink 2007. *Evaluating Research in Context: a Method for Comprehensive Assessment*. The Hague: COS.

Stokes, D E 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.

Stolker, C J M M 2003. 'Ja, geléérd zijn jullie wel!' Over de status van de rechtswetenschap. *Nederlands Juristenblad*, **15**, 776.

TU Delft 2009. *Gouden BTA Regels voor verslagjaar 2009*. Available at <https://intranet.tudelft.nl/live/pagina.jsp?id=f54c 6d15-df61-411b-b2fa-a1162f2e8910&lang=nl>, last accessed 28 October 2010.

Van der Voordt, T J M 1999. *Methodologie voor Bouwkunde*. Delft: Faculteit Bouwkunde, B-Nieuws.

Van Leeuwen, T N 2006. The application of bibliometric analyses in the evaluation of social science research: who benefits from it and why it is still feasible. *Scientometrics*, **83**(2), 363–374.

Van Raan, A F J, T N van Leeuwen, M S Visser *et al* 2010. Rivals to the crown: reply to Opthof and Leydesdorff. *Journal of Informetrics*, **4**(3), 431–435.

VSNU 2005. *Oordelen over rechten*. Rapport van de Commissie Voorbereiding Onderzoeksbeoordeling Rechtsgeleerdheid. Utrecht: VSNU.

VSNU 2007. *Naar prestatie-indicatoren voor rechtswetenschappelijk onderzoek*. Rapport van de Commissie Prestatie-Indicatoren en Ranking. Utrecht: VSNU.

Waltman, L, N J van Eck, T N van Leeuwen *et al* 2011. Towards a new crown indicator: some theoretical considerations. *Journal of Informetrics*, **5**(1), 37–47.

Whitley, R. 2000. *The Social and Intellectual Organizations of the Sciences*. Oxford: OUP.