in press
Research Evaluation

How accurately does output reflect the nature and design of transdisciplinary research programmes?

Elizabeth Koier* and Edwin Horlings*

*Rathenau Instituut, Science System Assessment Department, Anna van Saksenlaan 51, 2593 HW, The Hague, the Netherlands

Corresponding author: Elizabeth Koier, e.koier@rathenau.nl, tel. +31-70-3421542

Acknowledgements: We gratefully acknowledge the valuable feedback and suggestions of the two anonymous reviewers and of Jasper Deuten, Stefan de Jong, and Tjerk Wardenaar.

Funding: This work was supported by the Dutch national research programme Knowledge for Climate [grant number SSA01] (www.knowledgeforclimate.org). This work was part of a meta-project to monitor the organisation, effects, and impacts of the programme called *Comparative Monitoring of Knowledge for Climate*. The Knowledge for Climate research programme is co-financed by the Ministry of Infrastructure and the Environment.

Abstract:

Many of today's societal problems are wicked problems that require a new, transdisciplinary approach in which knowledge of scientists and stakeholders and from different disciplines is integrated. The evaluation of transdisciplinary science requires a multi-method approach. Bibliometric analysis is consistently among the methods in multi-method evaluations. We analyse the accuracy of bibliometric evidence for the evaluation of transdisciplinary research by examining two large climate adaptation research programmes in the Netherlands. The assessment of accuracy involves a comparison of different approaches to defining and measuring involvement, output, and quality. We draw three conclusions with regard to accuracy. First, scientific output covers a fairly high amount of the scientific activities of the programmes, though information on funding agencies is not yet sufficiently accurate to reconstruct a programme's output through the Web of Science. Second, scientific output does not accurately reflect the nature and design of the programmes. The Web of Science appears to underestimate locally oriented and practically oriented research, non-academic actors rarely co-author scientific publications, and the contributions of non-academic organisations to projects could not be recognised from author affiliations. Third, our exploration of two alternative reproducible metrics (non-scientific output and download statistics) shows that it is too early to introduce such metrics into evaluation practices. The research agenda for transdisciplinary output metrics should focus on the development of a common definition of transdisciplinary research output and a typology of non-scientific outputs, as well as a discussion and assessment of the relative value of such outputs for the integration of knowledge.

Keywords:

transdisciplinary research; evaluation; bibliometrics; large research programmes; Web of Science

1. Introduction

Governments around the world are using funding programmes and coordinating instruments to guide research towards themes and grand challenges that are relevant to society (Funtowicz and Ravetz 1993; Lyall and Fletcher 2013). Grand challenges are wicked problems that affect the interests of a multitude of stakeholders. Current thinking is that wicked problems require a new approach to the organisation of science, one that combines the strengths of different disciplines and that involves stakeholders in the development of solutions. Transdisciplinary science involves creating extended peer communities in which scientists and stakeholders have equal status (Meyer 2006; Spangenberg 2011) and integrate their knowledge, producing a wide range of knowledge outputs (Choi and Pak 2006; Rosenfield 1992), not all of which are scientific. A major challenge is that transdisciplinary science requires different criteria for quality assessment.

There is a growing body of literature on the evaluation of transdisciplinary science (Klein 2008; Wagner et al. 2011). This literature shows that the objective of the evaluation of transdisciplinary research is always to assess to what extent the integration of knowledge was achieved, although definitions vary, particularly with respect to the involvement of stakeholders (Klein 2008). No single metric exists with which to assess the integration of knowledge (Wagner et al. 2011). A multi-method approach is required (Stokols et al. 2003).

Bibliometric analysis is consistently used in multi-method approaches to evaluation. There is consensus that bibliometric analysis has its limitations, especially for programmes with societal impact as their main goal. Bibliometric data tend to measure outputs rather than outcomes, impacts, or process, and are biased towards the reward system of science (Pohl and Hadorn 2008). The validity of bibliometric evidence varies by scientific discipline and even in the evaluation of pure science programmes peer review is used to provide more reliable information (Geuna and Martin 2003). The purpose of this paper is not to propagate the use of bibliometrics in the evaluation of transdisciplinary research but rather to help ensure that they are properly understood and interpreted when they are used.

In this paper, we analyse the accuracy of bibliometric evidence for the evaluation of transdisciplinary research by examining in detail the participants and outputs of two large climate adaptation research programmes in the Netherlands. The aim of the paper is to determine whether and, if so, how accurately the design features of a transdisciplinary programme are reflected in its output. The issue is not whether bibliometrics can accurately recognise the design features of different modes of scientific research (such as pure science, interdisciplinary research, mode-2 research, or university-industry collaborations) or whether bibliometrics can more accurately recognise the design features of transdisciplinary research than of other modes of research.The assessment of accuracy involves a comparison of different approaches to defining and measuring involvement, output, and quality.

In section 2, we discuss the validity of the use of bibliometric and quantitative methods for evaluation of transdisciplinary programmes. In section 3 we present the data. In section 4 we discuss the results on the accuracy of publication data as source of information on transdisciplinary programmes. Our main conclusions and their implications are presented in section 5.

2. What do output metrics show for transdisciplinary science?

Transdisciplinary research programmes serve a variety of goals, such as network formation, capacity building, transdisciplinary cooperation, development of practical solutions for societal problems and scientific advancement (Hegger et al. 2012; Hessels et al. in press; Pohl 2008). Bibliometric analysis – looking at the volume, impact, and co-author networks in peer-reviewed scientific output in (international) journals – is among the conventional methods for evaluating the results of research programmes. Given the nature of transdisciplinary and problem-oriented science, what can output metrics show and what can they not?

There is a growing body of literature on the evaluation of interdisciplinary and transdisciplinary science. A review of the evaluation literature by Klein (2008) shows that there are three strands of literature on (1) interdisciplinary research, (2) transdisciplinary team science, and (3) transdisciplinary trans-sector, problem-oriented research with stakeholders in society. The second cluster is predominantly American, the third mostly European. In this paper, we focus on transdisciplinary research only and use the European interpretation of transdisciplinary research as trans-sector, problem-oriented research involving stakeholders as equal partners.

There is consensus on three issues regarding transdisciplinary science: its general definition, its diversity, and its principal outcomes. Most authors adhere to the definition of Rosenfield (1992) that focuses on the integration of knowledge by researchers and stakeholders from different fields, working on a common problem over an extended period of time, developing shared conceptual frameworks, skills, and goals, transcending disciplinary perspectives (Choi & Pak 2006). Especially where scientists work with stakeholders, transdisciplinary science involves a wide range of actors – each with their own goals, expectations, motivations, and value systems – who interact socially and cognitively (Klein 2008). The outcome of transdisciplinary science emerges from this interaction. The principal outcome of transdisciplinary science is the integration of heterogeneous knowledge from different disciplines and sectors. Both process and output can be highly diverse. Evaluation is consequently a major challenge.

Establishing whether knowledge integration has truly been achieved requires a broad, multi-method approach, including interviews, focus groups, surveys, and document analysis, as well as bibliometric analysis of output (Stokols et al. 2003). In a review of the literature on interdisciplinary research, Wagner et al. (2011) observe that the process of knowledge integration is more difficult to observe than its results. Also, they see a distinction between those who examine the social and cognitive processes of knowledge integration in teams and those who study interdisciplinarity as part of the internal dynamics of the knowledge system. This distinction is comparable to the micro and macro levels of analysis of Stokols et al. (2003). The validity of metrics depends on the perspective of the evaluator and the purpose of the evaluation: system (macro) versus teams (micro). The two types of evaluators use different methods and metrics. For evaluators on the system level, “a valid assessment of the interdisciplinarity of research must involve some indication of the degree or extent of knowledge integration that took place as the research was being conducted” (Wagner et al. 2011, p.3), which requires measures that “reflect the underlying social dynamics and/or motives of the researchers” (Wagner et al. 2011, p.12). The evaluators on the micro level are more likely to apply quantitative indicators that have been developed over time and are considered valid for the analysis of interdisciplinary and transdisciplinary output. Monitoring and evaluation require reproducible measures and indicators, but there is as yet no single metric that can serve as a proxy (Wagner et al. 2011). Although they focus on interdisciplinarity, the conclusions of Wagner et al. also apply to transdisciplinarity research.

Conventional bibliometric indicators are considered valid within specific approaches and for specific purposes. Bibliometric indicators favour the reward system of science (Dasgupta and David 1994; Merton 1957), the ‘home base’ of scientists (Pohl & Hirsch Hadorn 2008). Scientists and stakeholders have different incentives. Scientists expect to achieve new scientific insights and research outputs, while stakeholders favour changes in knowledge to support decision making (Walter et al. 2007). This is where bibliometric methods fall short.

Even within science, there is increasing critique on the (sole) use of bibliometric indicators for evaluation. Geuna and Martin (2003) conclude that the validity of bibliometric evidence cannot be applied to every university department and scientific discipline and peer review is needed to provide more reliable information. Schmoch et al. (2010) note that output is one-dimensional, whereas research is a multi-dimensional. A focus on research performance and excellence neglects other dimensions of performance. De Jong et al. (2011) acknowledge the disciplinary diversity of output – books in the humanities, journal articles in physics, conference proceedings in computer science, designs in architecture, legal advice in law – and propose developing discipline-specific and group-specific indicators linked to the mission and audiences of research groups. They propose a focus on productive interactions between scientists and their non-academic stakeholders rather than on visible impact. Probst et al. (2011) propose using profiles, combining different types of output, to measure the output of the social sciences and humanities. In addition, scholars produce different types of output at different stages of their life cycle (Braam and van den Besselaar 2010) and they work in several problem areas at the same time (Horlings and Gurney 2013), each of which has its own sources of funding, network and objectives. In other words, even within science a single metric is not sufficient and correct attribution of output to its source – essential in the evaluation of funding programmes – is not always straightforward.

It is clear that bibliometric analysis serves a very specific purpose. A full evaluation of transdisciplinary science requires a mixed-methods approach, including a broader set of reproducible metrics that does justice to its diversity. Klein (2008) proposes seven generic principles for the evaluation of transdisciplinary research. Evaluation needs to account for (1) variability of goals, (2) variability of criteria and indicators, (3) leveraging of integration, (4) interaction of social and cognitive factors in collaboration, (5) management, leadership, and coaching, (6) iteration in a comprehensive and transparent system, and (7) effectiveness and impact. Bibliometric analysis can inform and support evaluation according to Klein's principles. For example, the properties and contents of scientific output can reflect the variability of goals (principle 1), such as the production of new knowledge, the development of technical equipment, or the combination of methods from different fields. Traditional field-based quality indicators, such as numbers of patents, publications, and citations, are part of – but by themselves not sufficient for – measuring the variability of criteria and indicators (principle 2). The author lists of academic and non-academic outputs provide an indication of who is involved in the social process of knowledge production and can be seen as a proxy for communication and participation (principle 4, interaction of social and cognitive factors in collaboration). The question is not if bibliometrics is sufficiently accurate to substitute other, qualitative methods for the evaluation of transdisciplinary research programmes. We know that bibliometric analysis has its limitations and cannot provide sufficient information for a full application of Klein's principles. However, it is consistently among the methods used in evaluation and has the advantage of the persuasive power of quantitative analysis. The question is if bibliometric analysis – when applied to certain parts of an evaluation – can provide an accurate picture of the processes and diverse outputs of a transdisciplinary research programme.

3. Data

To assess the accuracy of programme-related data, we collected data on the participants, projects, and outputs of two Dutch climate programmes from the programme's own databases as well as from the Web of Science. In this section, we describe the two climate programmes that were studied as well as the datasets that have been used in the analysis.

3.1. Two climate programmes in the Netherlands

Climate is consistently among the themes that are targeted by transdisciplinary science. In the Horizon 2020 Framework Programme, the European Union has reserved €30 billion of the total budget of €72 billion to tackle Grand Challenges. Sustainability development, in particular climate action and resource efficiency, are at the core of Horizon 2020: an estimated 60 per cent of the total budget will be related to sustainable development and 35 per cent of the budget will be climate-related (European Commission 2011). Earlier, in 2009, the European Institute of Innovation and Technology set up a Knowledge and Innovation Community (KIC) to integrate education, research and innovation in the area of climate. In addition, most OECD countries have special programmes and institutes for climate adaptation and mitigation. Most of these are scientific, but quite a few involve collaboration with stakeholders.[1]

The two Dutch programmes studied in this paper focus on (local) adaptation to and mitigation of the effects of climate change. The projects funded by the two programmes were selected with a preference for consortia of universities, public research organizations, industry and (lower) governments. This means that although many projects in the programme were transdisciplinary, not all activities or projects involve stakeholders. The two programmes have a specific governance structure (a foundation) that sets them apart from universities and research councils.

The mission of the first programme *Climate changes Spatial Planning* (*Klimaat voor Ruimte*; KVR; 2004-2011) was to equip government and industry in the Netherlands with an operational knowledge infrastructure tailored to the relationship between (anthropogenic and natural) climate change, climate variability, and land use. The programme aimed to encourage ‘climate-responsible’ spatial planning by looking for innovative ways to fulfil economic, societal, and ecological demands for space that arise from multifunctional and flexible use of wet and dry spaces, both below and on the ground.[2] An ulterior goal of the programme was to create new network relations among researchers and stakeholders in the area of climate adaptation and mitigation. The total budget of the programme was €80 million consisting of a €40 million grant and €40 million in matched funding from the participants in the programme’s projects . The first projects were started in 2004. In 2007 the programme was extended until 2011. KVR was integrated with Knowledge for Climate, which began in 2008. Over time, the programme grew into a consortium of 155 partners including universities and public research organisations, ministries, provinces, water boards, environmental NGOs and engineering consultancies. The programme encouraged stakeholders from non-academic research institutes as well as governmental bodies and industry to collaborate with academic researchers (Hegger et al. 2012). Most projects were carried out by more than one institution and in many cases by a variety of academic and non-academic organizations.

The second programme *Knowledge for Climate* (*Kennis voor Klimaat*; KVK; 2008-2014) was designed to develop the knowledge and tools with which it is possible to

---

[1] Wardenaar (2012) identified 56 climate adaptation research programmes, 14 of which involve stakeholders.
[2] http://www.klimaatonderzoeknederland.nl/klimaatvoorruimte (Accessed 28 October 2013).

determine if the spatial and infrastructural investments of the next twenty years are climate-proof and, if necessary, to adapt them. The mission of the programme is to develop scientific and applied knowledge for a climate-proof spatial planning of the Netherlands and to create a sustainable knowledge infrastructure for dealing with climate change.[3] Government (national, provincial, local, water boards) and firms actively participated in agenda setting (van der Weijden et al. 2012; Wardenaar et al. under review). The programme has a total budget of €92 million, half of which consists of matching funds provided by programme participants and alternate sources of project funding (Spaapen and van Drooge 2011). KVK has two types of projects. In *thematic projects*, regular projects focus on scientific research. Funds were divided through open calls for proposals. In *hotspot projects*, researchers, governments, firms, and NGOs together develop and implement options for climate adaptation strategies. Their work involves articulating a demand for knowledge, knowledge development, and the application of knowledge. Hotspot projects are the core of stakeholder involvement in KVK.

3.2. Datasets

The activities supported by the programmes were administrated in various databases. The authors of this paper were granted complete access to data from the programme's administration. In this paper, we examine and compare the following datasets.

1. *Database of programme participants*: This database is maintained by the programme office for accountability purposes and contains full information on over 2,000 unique individuals, including their academic and professional titles, their institutional affiliations, and their participation in the projects of the two programmes. The database links programmes to projects and projects to participants and organisations. Our version contains all programme participants as per December 2012.
2. *Project details*: The programme office maintains a list of projects. Each project has a title and description, involves a number of participants, and has a budget and a formal start date and end date.
3. *Database of programme outputs*: The programme maintains a public database in which all project outputs can be uploaded by project leaders. There are various categories in the upload system, but there are no protocols as to what should and should not be added to the database. The output database contains all public communications varying from television interviews to peer-reviewed scientific papers. The programme encourages projects to upload their outputs, but the contents are added to the database at the discretion of the project leaders. Our version contains all reported programme outputs as per February 2013.

---

[3] http://knowledgeforclimate.climateresearchnetherlands.nl/programme/mission-and-approach (Accessed 28 October 2013).

4. *Download statistics*: The webmaster of the Knowledge for Climate programme provided access to the statistics of downloads from the programme database. The data include the total annual number of downloads of each individual item in the programme database in 2009, 2010, 2011, and 2012, and the monthly number of downloads of each individual item from November 2012 up to and including October 2013.
5. *Web of Science output*: Peer-reviewed scientific articles that are published in specific journals are indexed by one or more of the major scientific indexes, such as Thomson Reuters' Web of Science and Elsevier's Scopus. The indexes contain publication meta-data, including – since recently – information on funders. In this paper, we use the Web of Science to trace the output of the two programmes as stated in the database of programme output.

The datasets were linked in order to relate programmes to projects, projects to participants, participants to authors, authors to outputs, projects to outputs, and outputs to downloads. Within the programmes' databases, links between programmes, projects, participants, and outputs were embedded. Links between participants and authors were established by means of names and affiliations as well as – in ambiguous cases – using Google results and LinkedIn profiles.[4]

3.3. Relationships between data types

The ways in which a programme produces output and generates impact can be summarized as follows, although it is not necessarily linear.

- A *programme* funds *projects*, setting terms and conditions such as the requirement to obtain matching funds, and differentiating between projects with different objectives. In the two climate programmes, there is a difference between research-oriented projects, action-oriented projects and communication projects.
- *Projects* are carried out by individual *participants* who either organise their own participation or are delegated by their organisation. We may assume that senior scientists tend to organise their own participation and then attract PhD students; participants from firms, government agencies, and NGOs tend to be delegated by their organisation. Participants can be scientists or stakeholders.
- In projects, *participants* produce *output*. Output comprises peer-reviewed scientific articles in the journals index in the Web of Science, Scopus or other indexes or in non-indexed journals, other scientific papers and reports, policy reports and other grey

---

[4] If people had the same last name and the same initial(s) it may be that we occasionally linked names that belong to different persons. However, we expect this will concern a very small group of persons, considering the limited size of the research community, the relations between different data in the databases, and our detailed check on affiliations.

literature, tools, techniques, presentations, lectures, interviews, and other tangible forms of output. Output is generally classified as scientific or non-scientific.

- *Output* generates *outcomes* and *impact*. Outcomes and impact are also generated by intangible results, such as tacit knowledge, experience, and awareness, as well as through the mobility and social networks of participants and other productive interactions (Spaapen and van Drooge 2011). Outcomes involve changes in the attitudes and behaviour of persons and organisations involved in projects; impact refers to effects on the target problem. In the two climate programmes, impact is achieved when knowledge with respect to climate adaptation is applied and better prepares the Netherlands for climate change.

This process is not necessarily linear. Participants can initiate new projects, based on the results of prior work within the programme and using their first-hand information about calls for proposal and the programme's priorities and procedures. Through the social networks of participants, new organisations and individuals may be attracted into new projects. Outputs can steer the programme in new directions. Major achievements can attract new public and private funds that extend the programme's duration and create opportunities for new projects, new participants, and new outputs.

In the remainder of this paper we will examine the accuracy of current output indicators for the two climate programmes, investigate the usability of download statistics, a new and upcoming indicator, and assess the accuracy of links between data on participants, projects, and outputs.

4. Testing accuracy

The literature review in section 2 show that the accuracy of bibliometric indicators needs to be tested before they can be used. Also, additional sources of information are needed to give a more reliable indication of the output and impact of a programme. Many large programmes, like the climate programmes studied here, aim to develop practical solutions for wicked societal problems. To achieve this aim, they provide incentives to collaborate with stakeholders who have a practical agenda and encourage the production of a wider range of outputs, for example including tools, models, presentations, and policy studies. This means that we need to take into account both the scientific output and the non-scientific output of the programmes as well as the interactions between the different types of output.

In this section, we assess accuracy using the following questions:

1. Does scientific output accurately reflect the activities, nature and design of the programmes?
2. Do the programme and publication data accurately reflect the contribution of programme members to the performed research?

3. Are there alternative reproducible metrics that provide a more accurate picture of the non-scientific output of transdisciplinary science?

4.1. Scientific output

The first step in our accuracy test concerns scientific output, the key performance indicator in the evaluation of research programmes. We compare two sources: (1) the programme's own list of peer-reviewed scientific articles reported by project leaders, and (2) Web of Science (WoS) metadata for those articles in the list that were published in journals indexed by the WoS.

It is a well-known fact that the WoS does not provide full coverage of scientific output. Some disciplines (e.g. medicine, physics) are represented more completely than other disciplines (e.g. the humanities). The same is said about countries, where the WoS is assumed to contain a larger proportion of the total scientific output of some countries (e.g. the United States) than of other countries (e.g. China). However, Wagner and Wong (2012) found that the percentage of national venues for scientific output covered by the WoS for BRICS countries is similar to that of advanced countries (c. 3 per cent) and that researchers in BRICS countries have sufficient access to high-quality venues. In addition to disciplinary and national differences in coverage, the WoS – like other indexes – is continuously expanding its coverage, especially in specific areas (e.g. social sciences and humanities; conference proceedings), making longitudinal analysis more difficult.

In practice, in many disciplines incomplete coverage is not considered an obstacle to bibliometric evaluation as is shown by its abundant use. The main argument is that the WoS is considered an index of high-quality journals that publish knowledge of global importance. In many disciplines, evaluators feel they can afford to forego articles not covered by the WoS. This is definitely not true for the social sciences and humanities, where WoS coverage is relatively low and a large proportion of output is published in national languages, in the form of books, or in journals of local importance.

We compare the list of peer-reviewed scientific articles reported by project leaders with a list of articles in the programme database that were published in journals indexed by the WoS. What percentage of total output is indexed? And is the distribution of coverage random or biased in relation to the programmes' nature and design?

Table 1. Peer-reviewed publications in the programme output database according to Web of Science coverage

	Number of publications	Percentage of group	Percentage of the total number of publications
Total publications	480		100%
Found in the WoS	397	100%	83%
published in a Dutch language journal	0	0%	
local (spatial) planning and policy	19	5%	
Not found in the WoS	83	100%	17%
of which:			
published in a Dutch language journal not indexed by the WoS	19	23%	4%
published in an international, non-local journal not indexed by the WoS	21	25%	6%
published in an international, local (spatial planning and policy) journal not indexed by the WoS	9	11%	
other publications not covered by the WoS (e.g. books)	17	20%	4%
too recent to be found in the WoS (2012 or more recent)	2	2%	0%
unclear [a]	15	18%	3%

[a] The publication source is unclear in the programme database (n=2) or the publication source is a journal indexed by the WoS but the article was not found in the WoS (n=13).

We were able to find about 83 per cent of the peer-reviewed articles in the output database in the WoS, using both automatic retrieval and manual corrections. This is roughly comparable to the results of Ingwersen and Larsen (2007), who find that the WoS covers 79 per cent of the reported output of the Danish Strategic Environmental Research Program SMP. As can be seen in Table 1, 23% of the output that could not be found was published in Dutch language journals. None of the Dutch language journals in this dataset are indexed in the WoS. Papers published in international journals with a more practical or local orientation account for 11% of output not found in the WoS against 5% of output found in the WoS. Most of these are journals on law or practical (governmental) issues such as disaster management, spatial planning, and the effects of hazards on inhabitants. This is also compatible with the findings of Ingwersen and Larsen, who observe that “non-SCI journals tend to be broader practice-related international journals or magazines in Danish.”

A key assumption in the linked datasets is that outputs in the programme database were made possible at least in part by the financial support of one of the two programmes. This implies that we might use funder information to retrieve programme outputs. In many cases, a list of articles produced by the programme is not available. Constructing programme output by collecting all articles produced by programme participants during the programme’s duration leads to overattribution, since scientists tend to work on various projects simultaneously, using funds acquired from different sources (Horlings and Gurney 2013). Since August 2008, the WoS provides a second method of entry by extracting funder information from the acknowledgements in publications. As a result, funder information is increasingly used to assess programme outcomes (Costas and van Leeuwen 2012; Rigby 2011, 2013). The accuracy of this approach is less well known. We have tested to what extent the programmes' publications can be retrieved from the WoS using funder information.

Table 2. Retrieval of acknowledgements and funders by the Web of Science

	Number of publications	Percentage of the total number of publications
Total publications	221	
Acknowledgement was correctly retrieved	***142***	***64%***
–the programmes' funders were correctly identified	57	26%
–the programmes' funders were not correctly identified	26	12%
–funders do not include one of the programmes	59	27%
Acknowledgement was mistakenly not retrieved	***54***	***24%***
–funders include one of the programmes	34	15%
–funders do not include one of the programmes	20	9%
There is an acknowledgement but it does not contain funder information	***7***	***3%***
There is no acknowledgement	***18***	***8%***

Table 2 presents the results. The first observation is that researchers do not always acknowledge their funders. We found that about 53 per cent of the programmes' output from 2009 onwards contains an acknowledgement to one of the programmes. The second observation is that there are inaccuracies in the retrieval of funder information where researchers do acknowledge their funders. The WoS derives its funding information in two steps. First, it recognises and retrieves the acknowledgement. Then, it extracts individual funders from the acknowledgement text. We found that both steps contain inaccuracies. To give an example: One of the articles contains the following acknowledgement, which is recognized as such by the WoS:

> 'The authors would like to thank [...] and two anonymous reviewers for their valuable comments. This research was partly carried out in the framework of the Dutch National Research Programme "Knowledge for Climate" and "Climate Changes Spatial Planning" (http://www.climateresearchnetherlands.nl) and has been cofunded by The Netherlands Organisation for Scientific Research (NWO). The authors also thank their colleague [...] for his comments on an earlier version of this article.'

The funder information in the WoS on this article, however, only reads: "The Netherlands Organisation for Scientific Research (NWO)". The acknowledgement to the two climate programmes is not recognized as funder information.

The first step – recognition of the acknowledgement – is responsible for a loss of information on 24 per cent of the articles. The second step – extraction of individual funders where the acknowledgement was correctly retrieved – involves a loss of information on another 12 per cent of the articles. Of the 117 articles that do acknowledge funding by one of the two Dutch climate programmes (57+26+34), 60 articles (26+34, 51 percent) were not recognizable as such from the WoS funder list.

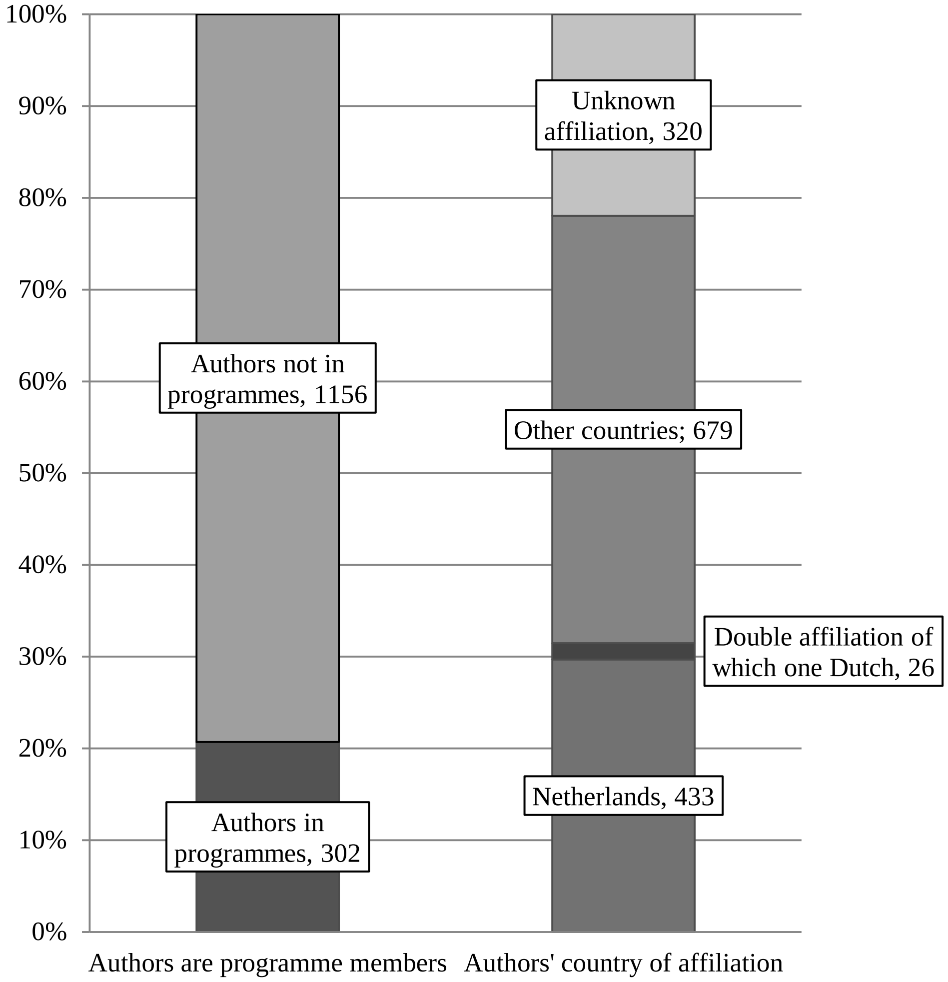
We can draw the following conclusions. Measuring the scientific output of climate programmes only on the basis of the Web of Science underestimates the output in practice-oriented areas, including locally oriented journals. This implies that we underrate performance with respect to one of the primary goals of the programmes: to produce practical solutions for local problems. We can also conclude that the WoS covers a high percentage of the programmes' scientific output, but that outputs that are closest to the practical, locally oriented nature of the programme are underrepresented. Also, funder information in the WoS is as yet not sufficiently accurate to reconstruct a programme's output. Programme funding is only acknowledged in about half of the publications, and where programme funding is acknowledged, the WoS extracts about half correctly. Part of the information loss may be attributed to the carelessness of researchers or to differences in publication practices with respect to acknowledgements. An alternative explanation is that some outputs in the programme database are not the result of the programmes. We will examine this particular phenomenon in closer detail in the next subsection.

4.2. The contribution of programme members to output

For the purposes of evaluation, accurate attribution of output is of vital importance. The first, general question on attribution is whether the papers assigned to the projects were written by authors who participate in the same projects. The involvement of stakeholders is a defining characteristic of transdisciplinary research programmes. The second question is, therefore, if the output of a programme reflects its nature and design, specifically with respect to stakeholder involvement.

Are project members co-authors of the papers assigned to the projects in which they participate? We find that in most articles (83 per cent) at least one author is indeed a member of the project to which the article was assigned. The remaining articles (17 per cent) fall into two groups. In the first group (5 per cent) the authors were members of other projects within one of the programmes than the project to which the article was assigned. The second group (12 per cent) contains articles of authors who were not project members themselves but who were colleagues of project members. Many of these mismatches occur when projects produce new equipment that is also used for research by others in the organisation who are not officially involved in the programme. This output has nonetheless been reported as project output.

Figure 1. Authors of programme output by membership of the programme and by country of affiliation



In Figure 1 we look at all authors in programme output distinguished by programme membership and by country of affiliation. The first bar shows that the number of authors who are not involved in the programme outnumber those who are involved by almost four to one. Even in a nationally oriented programme, there is a lot of collaboration outside programme boundaries. The second bar shows the country of affiliation of the authors. Since (almost) all programme members have a Dutch affiliation, we may conclude that many of the non-programme co-authors are international co-authors, although there is also a small group of Dutch authors who are not programme members. This means that the money allocated to a project is amplified by drawing in external co-authors from the network of the project members. Some of these may be stakeholders but they are more likely to be fellow scientists from all over the world. In short, co-author networks in output are a reflection of a variety of processes, not all of which link to the nature, design and activities of the programme.

Co-authorship by non-academic organisations has been used to evaluate patterns and trends in public-private research collaboration (Abramo et al. 2010; Tijssen 2011, 2012). While advertising the many benefits of this type of analysis, Tijssen (2012) does acknowledge that public-private co-authorships probably underestimate the true amount of collaboration between academics and firms. One of the main reasons is that publications do not function as rewards for non-academic participants. However, this does not prevent researchers and policy makers from drawing conclusions from public-private co-authorship analyses.

We have examined to what extent stakeholder involvement in the programmes and their projects is reflected in programme output. Since the Knowledge for Climate programme is still in progress, the analyses only relate to the Climate changes Spatial Planning programme. In 17 of the 33 CcSP projects with scientific output (52 per cent), all consortium partners are represented in the affiliations of authors of scientific output in the WoS. In the remaining 48 per cent, the division of the missing consortium partners is as follows: industry (33 per cent), PRO's (23 per cent), NGO's and universities (both 13 per cent) and governments (10 per cent).

The left side of table 3 shows the number of articles of which at least one author is affiliated to a type of organization. If we look at the same data from a project perspective (right side of the table), most of the projects (11 of 13) involving governments, firms and NGO's could not be recognised through author affiliations. This clearly shows the bias described by Tijssen (2012).

Table 3. Affiliations in project output

	Articles co-authored by at least one author affiliated to types of organisation N=304		Projects of which the consortium partners are recognisable from author affiliations in project output N=33		
type of organisation	Dutch and foreign affiliations	Dutch affiliations only	total number of projects involving organisations	projects not recognisable from author affiliations	percentage not recognisable
University	251	202	23	4	17%
PRO	193	155	26	8	31%
Government	6	5	3	2	66%
Firm	27	17	7	6	86%
NGO	6	0	3	3	100%

4.3. Alternative reproducible metrics

Transdisciplinary research programmes produce science to fulfil a societal mission. They produce more than (peer-reviewed) scientific articles and aim to generate an impact outside science. Results should be used to tackle societal problems, preferably by stakeholders. Programmes collect information that provide insight into the diversity and usage of their outputs. Can such information provide an alternative reproducible metric for the non-scientific output of a transdisciplinary research programme? We examine two types of programme data: output classified by the project leaders as 'non-scientific'[5] and the number

---

[5] Books, presentations, and posters may well be scientific. Many of the books in the programme database are reports, but a few books might be of a scientific nature. Here, we follow the classification of the programme database, which classifies books, presentations, and posters as non-scientific.

of downloads of the full programme output (i.e. both the output classified as ‘scientific’ and ‘non-scientific’) from the programme’s website.

The non-scientific output reported by projects and uploaded into the programme database consists of a wide range of items including factsheets, presentations, reports, articles for a broader public, newsletters, and press releases. These items represent the many ways in which projects – involving scientists as well as stakeholders – reach out to a range of audiences, using different methods and channels of communication. Table 3 shows the number of items per output type uploaded by projects into the programme’s output database. For every scientific item there are 3.6 non-scientific items of which 1.4 presentations, 0.7 reports, 0.4 media items, and 0.3 project factsheets. Figure 2 shows that non-scientific items dominate total output throughout the programme’s duration, especially from 2008 onwards.[6]

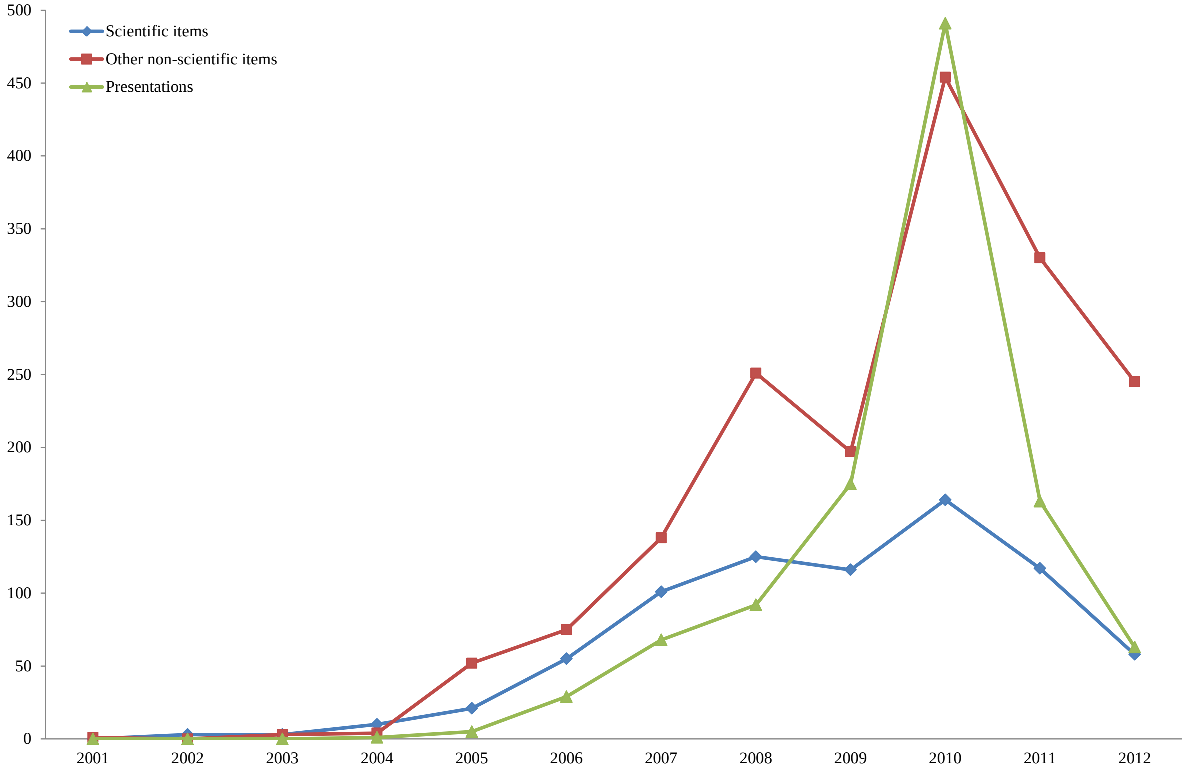
Table 4. Programme output by type according to the programme database, 2001-2013

Type of output	Count
Output classified by the project leaders as scientific	
Scientific papers	106
Peer-reviewed scientific papers	542
Proceedings	115
PhD theses	11
Output classified by the project leaders as non-scientific	
Audio	2
Books	43
Brochures	68
Final project reports	171
Media	322
Press releases	26
Popular articles about science	117
Posters	159
Presentations	1,087
Project factsheets	247
Project newsletters	81
Reports	515

---

[6] The peak in 2010 may be due to the overlap and merger of the two programmes as well as the fact that KVR was in its final year.

Figure 2. Development of the number of scientific and non-scientific items produced by the programmes according to the programme database, 2001-2012

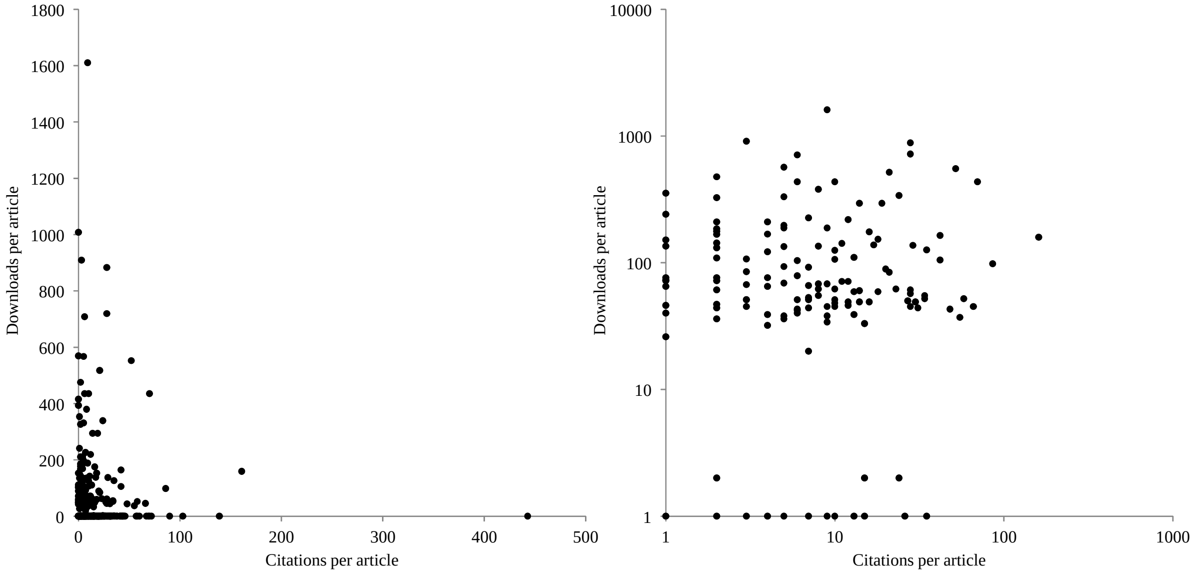


There are two problems associated with the programme's non-scientific output. First, the information on individual projects is inherently incomparable. The database contents depend mostly on self-reporting by project leaders who have different reporting habits. These differences affect the programme's aggregate output as well as the comparability of individual projects. One of the programme's communication projects illustrates how reporting habits affect the output database. The project reports several types of non-academic output, such as interviews with project members, but does not report that it has built several websites some of which are now the standard reference in newspapers and on twitter. Furthermore, some projects seem to have added all the (scientific) presentations that were held at conferences, whereas others only report published scientific papers. There is no standard for what needs to be added to the database and no incentive to report outputs that are not acknowledged in evaluations (van der Weijden et al. 2012).

The second problem is that there is no common definition or consensus understanding of the output of transdisciplinary research. This may account for the differences in reported outputs. Since the individual types of output are not comparable, it is also impossible to produce an aggregate measurement of non-scientific or societal output. How should we add up or weight presentations, newsletters, reports, and interviews? In pure science programmes, there is at least agreement on the unit of output, namely articles, preferably in high-impact journals, papers in the proceedings of high-status conferences, and books published with prestigious publishers. Transdisciplinary programmes geared towards societal problems aim to produce a wider range of outputs than pure science programmes. Also, non-scientific outputs are more than token contributions to society. They are of particular value to the stakeholders involved in the programme and deserve to be included in any measurement of output.

One possible solution to the incomparability of different items of non-scientific output is to use the statistics on annual downloads per item. If we consider downloads as the equivalent of citations, they may indicate the value of an item relative to other items. The use of download statistics and other similar indicators is propagated by a movement to produce alternative impact measures or altmetrics. Priem, Taraborelli, Groth, and Neylon (2010) have set the stage, calling for the development of (web-based) indicators that provide a more subtle picture of the many ways in which new scientific findings have an impact (see also Priem et al. 2012). Altmetrics include evidence of the sharing of datasets, code, and experimental designs and downloading, blogging, commenting, linking, and bookmarking of publications as well as passages from publications (so-called ‘nanopublications’). The general idea is that alternative metrics show how scientific results are actually used by different audiences in different contexts, other than by science itself but also including scientists themselves, for example in blogs (Groth and Gurney 2010).

Figure 3. Linear (left) and logarithmic (right) relationship between total citation impact and downloads in 2009-2012 of peer-reviewed scientific articles in the programme database



Theoretically, the number of downloads is a proxy for usage or impact and can be considered the equivalent of the number of citations to scientific articles. This has been tested with varying results (Guerrero-Bote and Moya-Anegón 2014; Shuai et al. 2012). Figure 3 reveals that in our case there is no relationship – either linear or logarithmic – between the total number of citations received and the number of times an item was downloaded from the programme database. There are two possible explanations: either the data are flawed or citations and downloads reflect different processes.

The download statistics of the programmes are inherently flawed for two reasons. The first reason is that items in the database can be downloaded from multiple sources. Citation counts are generally derived from a single source (e.g. the Web of Science or Scopus). These sources may not cover all of science, but it is clear which journals are included and each citation is measured only once. Download statistics provided by journals using an embargo period may be comparable in this respect, becauseduring this period a scientific article is presumably only available through the publisher’s website. This is makes download statistics from these websites -if they are provided at all by the publisher- during this period more reliable. However, after the embargo period, when the download counts from the programme start, this is no longer given.

In this particular case, and in many similar programmes, downloads refer not only to scientific papers which are downloadable from several sources after the embargo period, but also to a much wider set of outputs that, in some cases cannot be retrieved systematically from other sources. DBecause the website of the programme is accessible to anyone, downloading is done by a much wider audience, including stakeholders who generally have no access to (or interest in) scientific journals. Download statistics are consequently incomplete and have an unknown bias, given that some items are only available in the programme database while other items can be downloaded from many different sources.

The second reason is that download data as provided by the programme are not certain to be unique downloads contrary to citations. The download data of the Dutch climate programmes do not distinguish individuals or IP addresses on an item level. Even when they do, universities, companies, and other organisations often have a single IP address for all computers in their network. Consequently, it is possible for one individual to have been responsible for several downloads, for instance because it is an easy way to find the final version of one’s own paper or that of a colleague. As long as this metric is not used as an indicator or to make rankings, it is an interesting indication of the use of project results, especially since the programme database provides information on the usage of non-scientific outputs that is not available anywhere else. However, it is not suited for comparative evaluation.

Even if we were to overcome the abovementioned flaws in the data, downloads and citations measure different aspects of usage and impact, as was shown by Shuai et al. (2012), although Guerrero-Bote and Moya-Anegón (2014) found indications that in the period directly after publication there seems to be a correlation. Non-academic actors may use the programme outputs they have downloaded but are much less likely to cite them. Also, not all non-scientific items can be downloaded. Some items concern links to other websites and following a hyperlink does not count as a download. This means that – as it stands – download statistics from programme websites cannot serve as a quality indicator for non-scientific output.

5. Conclusions and discussion

Transdisciplinary science is on the rise and its evaluation requires a mix of methods. The diversity of transdisciplinary programmes and projects is bound to make evaluation a custom task, particularly in the absence of common definitions and standards of measurement. And yet, this is precisely what will be needed, should transdisciplinary science gain prominence.

In this paper we have examined the accuracy of one particular method from the mix, that of bibliometrics. Using all available information on the output of and participation in two large climate adaptation research programmes in the Netherlands, we have assessed the accuracy of the information on output and its relation to the programmes' organisation.

The first and most obvious conclusion is that bibliometrics is no substitute for the in-depth, multi-method analysis of knowledge integration in transdisciplinary science. It is, however, a very useful tool that can produce standardised and reproducible metrics with which the results of other quantitative and qualitative analyses can be compared and validated.

With regard to accuracy, we can draw three conclusions. First, the Web of Science (WoS) covers a high percentage of the scientific activities of the programmes: 83 per cent of scientific publications could be retrieved from the Web of Science. The recently added information on funding agencies is not yet sufficiently accurate to reconstruct a programme's output through the Web of Science. For now it is advisable to rely on the programme's own databases.

Second, scientific output does not accurately reflect the nature and design of the programmes in that the Web of Science appears to contain less locally oriented and practically oriented research. Non-academic actors – stakeholders – do not always co-author scientific publications and the contributions of non-academic organisations to projects could not be recognised from author affiliations.

Third, alternative reproducible metrics, such as downloads, social media references, blog references, and hyperlinks, are seen as a promising expansion on current citation-based impact indicators. Our exploration of two such metrics for non-scientific output– download statistics and the number of non-scientific output items – shows that it is too early to introduce such metrics into evaluation practices. There is no common understanding of the relative importance of different non-scientific outputs nor a standard for what needs to be reported. There is currently also no incentive for scientists to report outputs that are not acknowledged in evaluations. Downloads statistics on non-scientific ouput are not accurate enough for comparative evaluation. They are incomplete, most likely biased, and can be manipulated. Successful implementation of altmetrics for the evaluation of the non-scientific output of transdisciplinary science will require a degree of regulation and standardisation across the sciences. Standardisation may be useful for those who measure. It does not, however, solve the problems with accuracy identified in this paper. More than standardised output metrics, evaluation requires a narrative of impact and usage: how do projects and programmes generate an impact beyond science and how was that impact achieved, e.g. using various types of output?

The research agenda for transdisciplinary output metrics should discuss and assess the relative value of the range of transdisciplinary research outputs for the integration of knowledge and the achievement of societal impact. These tasks require a better understanding of the way in which programmes and projects integrate knowledge and generate impact through various outputs, processes, communication channels, and methods of stakeholder involvement. For example, the relative value of an item of non-scientific ouptut (e.g. a newsletter or a presentation) can be conceptualised as an impact that may be generated (effect) or a channel through which the output reaches an audience (productive interaction).

Another unresolved issue concerns the role of non-academic project participants as authors of various outputs (e.g. policy reports), as members of advisory boards and user groups, as workshop participants, etcetera. The position of non-academic actors in collaboration networks is a crucial addition to the information needed to evaluate how effectively a transdisciplinary programme involves stakeholders, disseminates knowledge, and produces impact. Finally, a key question in the evaluation of instruments that combine scientific and societal goals, such as innovation or climate adaptation, is whether there is a trade-off between societal and scientific achievement. Does excellence in one area go at the expense of excellence in the other area? This has been studied for patents (Breschi et al. 2008; Geuna and Nesta 2006; Meyer 2006) and it seems worthwhile to do the same for transdisciplinary science.

Evaluation requires solid information and it is in the interest of programmes to ensure that such information exists. Programme offices should make funding conditional upon funding acknowledgement and full reporting of all outputs to be collected and stored in a systematic manner. Ideally, when reporting outputs, project leaders should use a standard format that reflects a common understanding of non-scientific or societal output, keeping in mind that this understanding may vary by scientific discipline and societal problem area. Transdisciplinary programmes should explicitly include societal impact in their own evaluations, considering that this is their value added.

References

Abramo, G., D'Angelo, C. A., and Solazzi, M. (2010) 'Assessing public-private research collaboration: is it possible to compare university performance?', *Scientometrics*, 84 (1), 173-97.

Braam, R. and van den Besselaar, P. (2010) 'Life cycles of research groups: the case of CWTS', *Research Evaluation*, 19 (3), 173-84.

Breschi, S., Lissoni, F., and Montobbio, F. (2008) ‘University patenting and scientific productivity: a quantitative study of Italian academic inventors’, *European Management* Review, 5 (2), 91-109.

Choi, B. C. K. and Pak, A. W. P. (2006) ‘Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness’, *Clinical and Investigative Medicine*, 29 (6), 351-64.

Costas, R. and van Leeuwen, T. N. (2012) ‘Approaching the "reward triangle": General analysis of the presence of funding acknowledgments and "peer interactive communication" in scientific publications’, *Journal of the American Society for Information Science and Technology*, 63 (8), 1647-61.

Dasgupta, Partha and David, Paul A. (1994) ‘Toward a new economics of science’, *Research Policy*, 23 (5), 487-521.

De Jong, S., et al. (2011) ‘Evaluation of research in context: an approach and two cases’, *Research Evaluation*, 20 (1), 61-72.

European Commission (2011) ‘Horizon 2020 - The Framework Programme for Research and Innovation’, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (COM(2011) 808 final; Brussels).

Funtowicz, S. O. and Ravetz, J. R. (1993) ‘Science for the post-normal age’, *Futures*, 25 (7), 739-55.

Geuna, A. and Martin, B. R. (2003) ‘University research evaluation and funding: An international comparison’, *Minerva*, 41 (4), 277-304.

Geuna, A. and Nesta, L. J. J. (2006) ‘University patenting and its effects on academic research: The emerging European evidence’, *Research Policy*, 35 (6), 790-807.

Groth, Paul and Gurney, Thomas (2010) ‘Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study’, *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*.

Guerrero-Bote, VicenteP and Moya-Anegón, Félix (2014), 'Relationship between downloads and citations at journal and paper levels, and the influence of language', *Scientometrics*, 1-23.

Hegger, Dries, et al. (2012) ‘Conceptualising joint knowledge production in regional climate change adaptation projects: success conditions and levers for action’, *Environmental Science* & Policy, 18, 52-65.

Hessels, L.K. , et al. (in press: *Research Evaluation*) ‘The role of knowledge users in public-private research programs: an evaluation challenge’.

Horlings, E. and Gurney, T. (2013) ‘Search strategies along the academic lifecycle’, *Scientometrics*, 94 (3), 1137-60.

Ingwersen, P. and Larsen, B. (2007) ‘Evaluation of strategic research programs: the case of Danish environmental research 1993-2002’, *Research Evaluation*, 16 (1), 47-57.

Klein, J. T. (2008) ‘Evaluation of interdisciplinary and transdisciplinary research - A literature review’, *American Journal of Preventive Medicine*, 35 (2), S116-S23.

Lyall, C. and Fletcher, I. (2013) ‘Experiments in interdisciplinary capacity-building: The successes and challenges of large-scale interdisciplinary investments’, *Science and Public Policy*, 40 (1), 1-7.

Merton, R. K. (1957) ‘Priorities in scientific discovery - a chapter in the sociology of science’, *American Sociological Review*, 22 (6), 635-59.

Meyer, M. (2006) ‘Are patenting scientists the better scholars? An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology’, *Research Policy*, 35 (10), 1646-62.

Pohl, C. (2008) ‘From science to policy through transdisciplinary research’, *Environmental Science & Policy*, 11 (1), 46-53.

Pohl, C. and Hadorn, G. H. (2008) ‘Methodological challenges of transdisciplinary research’, *Natures Sciences Sociétés*, 16, 111-21.

Priem, J., Groth, P., and Taraborelli, D. (2012) ‘The Altmetrics Collection’, *PLOS One*, 7 (11).

Priem, J., et al. (2010) ‘Alt-metrics: a manifesto’ (http://altmetrics.org/manifesto/).

Probst, C., et al. (2011) ‘Profiles and beyond: constructing consensus on measuring research output in communication sciences’, *Research Evaluation*, 20 (1), 73-88.

Rigby, J. (2011) ‘Systematic grant and funding body acknowledgement data for publications: new dimensions and new controversies for research policy and evaluation’, *Research Evaluation*, 20 (5), 365-75.

Rigby, J. (2013) ‘Looking for the impact of peer review: does count of funding acknowledgements really predict research impact?’, *Scientometrics*, 94 (1), 57-73.

Rosenfield, P. L. (1992) ‘The potential of transdisciplinary research for sustaining and extending linkages between the health and social-sciences’, *Social Science & Medicine*, 35 (11), 1343-57.

Schmoch, U., et al. (2010) ‘How to use indicators to measure scientific performance: a balanced approach’, *Research Evaluation*, 19 (1), 2-18.

Shuai, Xin, Pepe, Alberto, and Bollen, Johan (2012), 'How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations', *PLoS ONE*, 7 (11), e47523.

Spaapen, J. and van Drooge, L. (2011) ‘Introducing 'productive interactions' in social impact assessment’, *Research Evaluation*, 20 (3), 211-18.

Spangenberg, J.H. (2011) ‘Sustainability science: a review, an analysis and some empirical lessons’, *Environmental Conservation*, 38 (3), 275-87.

Stokols, Daniel, et al. (2003) ‘Evaluating transdisciplinary science’, *Nicotine & Tobacco Research*, 5 (Supplement 1), S21-S39.

Tijssen, R. J. W. (2011) ‘ Joint Research Publications: A Performance Indicator of University-industry Collaboration’, *Evaluation in Higher Education*, 12, 19-40.

Tijssen, R. J. W. (2012) ‘Co-authored research publications and strategic analysis of public-private collaboration’, *Research Evaluation*, 21 (3), 204-15.

Van der Weijden, I., Verbree, M., and van den Besseaar, P. (2012) ‘From bench to bedside: The societal orientation of research leaders: The case of biomedical and health research in the Netherlands’, *Science and Public Policy*, 39 (3), 285-303.

Wagner, C. S. and Wong, S. K. (2012) ‘Unseen science? Representation of BRICs in global science’, *Scientometrics*, 90 (3), 1001-13.

Wagner, C. S., et al. (2011) ‘Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature’, *Journal of Informetrics*, 5 (1), 14-26.

Walter, Alexander I., et al. (2007) ‘Measuring societal effects of transdisciplinary research projects: Design and application of an evaluation method’, *Evaluation and Program Planning*, 30 (4), 325-38.

Wardenaar, T., Horlings, E., and Besselaar, P. van den (under review: *Environmental Science & Policy*) ‘Development of a typology of stakeholder roles in challenge-driven research programs and its application to climate adaptation programs in the US, Germany and the Netherlands’.