

# Evaluation and monitoring of transdisciplinary collaborations

Leonie van Drooge<sup>1</sup>  · Jack Spaapen<sup>2</sup>

Published online: 11 July 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** In this paper we focus on the governance, in particular evaluation and monitoring, of the growing number of transdisciplinary collaborations (TDC's). Researchers and a variety of stakeholders collaborate in such TDC's, the purpose of which is to address societal challenges, like renewable energy, healthy aging or better language teaching in schools. Commonly used practices for evaluation of scientific research (accountability, rankings and benchmarking, dedicated to scientific excellence) do not fit the goals of TDC's. A bottom up or stakeholder oriented approach is better suited; one that stimulates mutual learning as well as the development of socially robust knowledge. We introduce the participatory impact pathways analysis (PIPA), as a method that suits the requirements. It has been developed in the context of development research. Two crucial features are the involvement of stakeholders from the start, and the joint development of a theory of change. This narrates what one wants to achieve and how that will be achieved. From this, stakeholders construct a logical frame that serves as a source for indicators. These indicators enable monitoring *ex durante*, during the TDC. We present evidence of the use of PIPA for a TDC. From this empirical evidence a number of issues with regard to evaluation, monitoring and indicators can be identified that require attention. Most prominent is the change of function of indicators. Instead of looking back and a focus on past performance, indicators look forward, in the short, intermediate and more distant future.

**Keywords** Transdisciplinary collaboration · Evaluation · Research governance · Monitoring · Stakeholders · Theory of change · PIPA

**JEL Classification** O35 Social Innovation · O38 Government Policy

---

✉ Leonie van Drooge  
l.vandrooge@rathenau.nl

Jack Spaapen  
jack.spaapen@knav.nl

<sup>1</sup> Rathenau Institute, Anna van Saksenlaan 51, 2593 HW The Hague, The Netherlands

<sup>2</sup> Royal Netherlands Academy of Arts and Sciences, Kloveniersburgwal 29, 1011 JV Amsterdam, The Netherlands

## 1 Introduction

Scientific research, societal and industrial innovation and government policy increasingly are getting intertwined in transdisciplinary networks or consortia, TDC's. Whether it is at the supra-national level, for example the EU H2020 framework program, the national level, for example the Dutch top sector policy, or in mission oriented research institutes such as the French INRA, many governmental and institutional measures are targeting the enhancement of collaboration between science, society, industry and policy. The overarching goal of these programs is to focus (academic) research on those sectors or fields that are deemed vital for the economy, and/or on issues that are politically important or controversial. Examples of such important topics can also be found at the global level, for example climate change or migration, and in other regions of the world, for example the ASEAN plans of Action for S&T in South East Asia. One consequence is that public investment schemes are shifting from a main focus on fundamental and unbound research towards more applied research; or at least towards research that is taking the context of application into account, referred to as mode 2 research by Michael Gibbons and co-authors (Gibbons et al. 1994; Nowotny et al. 2001).

As a result, traditional forms of research evaluation no longer suffice. Research is part of a transdisciplinary consortium, and the context of application is relevant. One-size-fits-all indicators are not adequate to evaluate research in the context of a TDC; instead indicators are needed that suit each specific context and TDC. These can be quantitative as well as qualitative. Scientific excellence is important but not the only goal of TDC's; societal change is on a par. Commonly used indicators that relate to scientific aspects only, do not cover many of the aspects of a TDC. Therefore, a range of indicators is needed, that together describe the complexities of TDC's. The challenge for evaluation of TDC's is major because not only are different indicators needed, they also have to be attuned somehow in a meaningful way. An incremental change, such as an extra indicator relating to societal aspects, is not enough. A different view on and approach to evaluation is needed, one where the focus is on collaboration, and on reaching the goals of the TDC.

In this paper we review the changing policy and funding context (Sect. 2), the consequences of that process for the governance (Sect. 3) and evaluation (Sect. 4) of publicly funded research. We will zoom in on the significantly different requirements for evaluation of TDC's. We then present (Sect. 5) an alternative approach to evaluation—participatory impact pathways analysis (PIPA)—that is more adequate to serve research and innovation processes in a transdisciplinary context. The function of evaluation in PIPA is not so much that of an instrument for accountability, but rather a way to improve the research and innovation process, a learning tool. The application of PIPA provides empirical evidence (Sect. 6). It illustrates the outlines of a monitoring framework and provides insight into the range of indicators that can be developed. We conclude (Sect. 7) that, given that there is indeed a fundamental change in the context of research, evaluation methods, criteria and indicators, should change in a fundamental way too.

## 2 Changing policy context: the quest for relevance

The emergence of transdisciplinary forms of collaboration (TDC's) as an important form in research and innovation ventures, fits within a historical development. Research and innovation has become increasingly formalized to serve public policy goals at national and

supranational levels (Freeman and Soete 1997). The policy demand for research to address societal challenges dates back to the 19<sup>th</sup> century, when the universities became research universities with sometimes close relations to the government (Clark 2006). The relationship between government and academia changes considerably during and after the Second World War. Large public labs are set up in strategic research areas, with the Manhattan project as the prime example. The 1945 Vannevar Bush report “Science—the endless frontier” (Bush 1945) and the following establishment of the National Science Foundation can be seen as the birth of modern public science policy. The report aims at marrying the interests of fundamental science (unfettered) and political demand (societal relevance). It refers to the successes of the wartime investment in science. These delivered not only the atom bomb as a result of the Manhattan project, but also radar and penicillin (Pielke 2010). The report finds a captive audience.

In the 1970s governments start to demand more relevance from scientific efforts. Various strategic research programs are launched, for instance in the fields of micro-electronics and ICT. In the 1980s, many European governments start to develop policies for academic science that depend on conditions for relevance, like the so-called conditional finance system in the Netherlands (Blume and Spaapen 1988). Furthermore, industry-oriented research programs are introduced in many European countries. New arrangements are introduced for research aiming at industrial relevance, by setting up Cooperative Research Centres as new intermediary organisations (Gray 2011; Van der Veen et al. 2005). And in this century a number of national and European arrangements are being set up to support research oriented at global challenges (EC 2017a), such as Joint Programming Initiatives (EC 2017b) and Knowledge and Innovation Communities. At every turn, new collaborative arrangements are introduced into the research system, with the intention to enable and stimulate a joint academic and industrial effort (Stirling 2006; Hessels et al. 2009; Rip 2004). Gradually, the research system has become increasingly heterogeneous, with the quest for relevance as a strong driver.

Changes in the demands of public policy vis-à-vis academic research have led to changes in the requirements in many funding instruments. Even in programs and policies that are dedicated to science excellence, criteria are introduced that relate to societal relevance. In the US, NSF introduces broader impacts, in the Netherlands, research council NWO introduces knowledge utilization and in the UK, impact case studies are required for the Research Excellence Framework. A different example from the Netherlands, research council NWO is now obligated by the government to outlay half of its budget to research proposals that combine academic and industrial research. Interestingly enough, the main driver behind this policy is the ministry of economic affairs, and thus not the ministry of science and education. We see other changes as well at the level of research funders, that include new funding objectives (aimed at developing concrete solutions for societal challenges), new (joint) funding arrangements (e.g. crowdfunding), the entrance of new research funders (e.g. the Bill and Melinda Gates Foundation) and new review systems (e.g. extended peer review).

Through these developments, academic research finds itself increasingly in a context with strong demands from partners inside and outside academia. Moreover, many funding schemes are no longer targeting individual applications from researchers, but instead large collaborations between a variation of stakeholders aiming at solving major issues in society. When designing these large transdisciplinary research and innovation enterprises (TDC's), researchers together with other professionals have to address a wide array of issues that transcend the traditional scientific considerations.

TDC's aggravate an inherent tension within the academic system between the demand of being 'excellent' in the academic community and the demand of being relevant for society. While many researchers are motivated to contribute to societal issues (Lam 2011), current academic culture is at the same time dedicated to stimulating scientific excellence. Articles, citations and other metrics are what counts in the first place (Benedictus and Miedema 2016). Societal impact is of lesser importance in most current reward systems. The third mission of universities is perceived as a new and sometimes inappropriate task, with which scientists struggle (De Jong et al. 2016; Hessels 2010).

The ideas presented in the Bush report, in particular the idea that basic research should be performed without looking at societal impact, still influences academic culture, in the sense that it creates an artificial distinction between basic and applied research. However, there is a growing number of academics that leave behind the linear view of Vannevar Bush. Sarewitz (2016) calls the Bush doctrine "a beautiful lie". Sarewitz recognizes the fact that scientists these days are used to relating their fundamental research to relevant and urgent societal matters, but is afraid that it is often just lip-service. He quotes breast cancer spokeswoman Fran Visco, who called for action from the researchers in their labs: "at some point, you really have to save a life".

The Bush doctrine first started to wane in the 1960s and 1970s when it gradually became clear that science alone is not enough to conquer the major problems in society, and certainly not science as an pure academic endeavor. Nelson, in his famous book *The moon and the Ghetto* (Nelson 1977) argues that there are no clear paths to a solution regarding education for ghetto kids, whereas there was a clear path, in hindsight, for the development of the spacecraft Apollo. The heart of the problem, according to Nelson, is that there are political constraints and, above all, societal shortcomings. Society lacks know-how. Solving societal issues differs from an engineering challenge.

While we agree with Nelson's view that there are differences between an engineering challenge and a societal issue like education, we would argue that engineering challenges also are dependent on a societal context in which non-technical issues play a role (Nowotny et al. 2001). In fact there are many examples of engineering challenges that led to a lot of debate and controversies from both academics and non-academics (Blanckesteijn et al. 2014; Hoppe 2010).

Thus, we see an upcoming need for academics to address societal demand, which manifests itself in a changing funding culture and in a changing context to conduct research. Regarding evaluation, this leads to tensions between traditional academic instruments mainly based on measuring contributions in the scientific literature, and the quest for new instruments that can value the contribution of scientific research to societal issues. The introduction of TDC's in the research system requires a different view on the governance of such transdisciplinary collaborations.

### 3 Governance of TDC's

A main challenge for the governance of TDC's is the changing relationship between the funder and the research organization. Traditionally, the (national) government was the primary commissioner of academic research; in TDC's there often is a consortium like structure in which different funders participate. Instead of a top down situation in which it is clear who is accountable to whom (the researcher to the government that provides tax payers' money), in the new context, a more distributed situation develops in which the

interests of different participants ('stakeholders') co-exist. Kuhlmann and Rip (2014), reviewing the Grand Societal Challenges of the EU Horizon 2020 program, discuss the changes needed for these new ways of research, claim that new arrangements are needed for the governance of TDC's. They call for a more tentative form of governance, taking into account the dynamic process of TDC's, and refer to this as the 'challenge' of addressing the Grand Challenges. A similar lesson from an analysis of research projects addressing global challenges, is that ideally the governance structure should function as a 'learning system' (OECD 2012). Conventional arrangements are not suitable, since in TDC's a broad range of stakeholders is involved in different ways in the various stages of research (Kloet et al. 2013). This is substantively different from traditional forms of research governance, where stakeholders are not at all involved, or at a distance.

Thus, a prime challenge for governance of TDC's is to provide avenues for broad stakeholder engagement in the performance, governance and evaluation of TDC's. This implies that stakeholders have to be involved from the beginning of a TDC, and consequently are involved in the development of the agenda of the project, and in allocation and evaluation decisions. It also has consequences at the level of funders (ministries, research councils, supranational funders such as the EC in Europe). Policies are needed that serve these more diversified networks or ecosystems in which responsibilities and accountability are distributed between a variety of stakeholders. This will have consequences for the ways in which R&D budgets are allocated towards the TDC's that address societal challenges (Edler and Kuhlmann 2008). Moreover, at the level of intermediary organisations, such as research funding agencies, there is a need for new competencies for joint working practices, agenda-setting, programming, funding and evaluating of TDC's.

At the level of the 'classic' research organisations (universities, research institutes) adaptations are needed to enable and stimulate transdisciplinary research. Human resource management and evaluative schemes currently provide little rewards for broader impacts and public engagement. However, there are signs of a turning tide, see for example Van den Akker and Spaapen (2017).

Thus, the governance challenge of TDC is multifaceted and plays out at multiple levels simultaneously. In this paper, we focus on one aspect that is new ways of assessing the quality and relevance of the work conducted in TDC's.

#### **4 Effects on the evaluation of research and innovation**

Research evaluation is traditionally concerned with the appreciation of academic quality, and consequently with the output of academic staff in the scientific literature. Common goals of such evaluations are accountability for public funds (ex post) and underpinning of decisions about resource allocation (ex ante).

When it comes to TDC's, researchers, funders and other actors are still in need of insight into research quality. But research quality needs to be viewed in a broader context, a context that has been referred to by Nowotny et al. (2001) as the context of application. They call the knowledge needed for this: socially robust knowledge. This refers to the ways in which problems are perceived, defined, and prioritized by the stakeholders. It has implications for the ways in which scientific activities are organized (Nowotny et al. 2001: 117). Evaluation in this context therefor needs to include organizational aspects, such as expectations and assumptions of stakeholders, their knowledge needs and requirements.

So, while most current evaluation practices have evolved around values such as scientific excellence and management accountability (Whitley and Gläser 2007), it is clear that they are not sufficient to cover TDC's. We follow up on claims that changes in research evaluation are necessary to facilitate a change towards more interdisciplinary, application-oriented or responsible science (Hemlin and Rasmussen 2006; Wissema 2009). A range of approaches have been developed that address societal impact or relevance of scientific research (see for instance Spaapen and Van Drooge 2011; Joly et al. 2015). But evaluation of TDC's requires more substantial changes. The changes are similar to the changes needed for the evaluation of Responsible Research and Evaluation, RRI. An EU expert group that was to develop indicators for the evaluation of RRI, concluded that RRI, being a dynamic and multifaceted concept, would not benefit from a fixed set of indicators. It was rather in need of a toolbox of quantitative and qualitative indicators. The expert group concluded that the assessment of RRI required both indicators of the process and the outcome and impact of research and innovation. The indicators should support the learning process of the actors and organizations involved (Expert Group on Policy Indicators for RRI 2015).

This implies a new notion of accountability: rather than being held accountable for one's performance *ex post*, researchers incorporate their responsibility to society in their primary activities by involving stakeholders in the research process. This implies a participatory and distributed approach to evaluation in which stakeholders are empowered and committed. And although TDC's and RRI are not the same, the governance challenges are similar.

## 5 Evaluation of TDC's: theory of change and participatory impact pathways analysis

Before we elaborate on new ways of evaluation for TDC's, we highlight an important conceptual distinction between two main functions of evaluation: (1) evaluations primarily conducted to account for (summative evaluations) and (2) evaluations that aim at mutual learning and improving (formative evaluations) (Scriven 1991, 1996). The first function, where accountability is the prime motive, became popular in the wake of the New Public Management. The evaluation context is usually characterised by a unilateral relationship between a funding actor (a government, a research council, a university) and a research entity. Regarding the second function, where learning is the prime motive for evaluation, the focus is on the variegated contexts in which research and innovation takes place. This regards networks of multiple partners who, in more or less stable structures, work together aiming to solve a joint problem or question. Evaluation in such networks is a joint responsibility and will benefit from procedures in which different stakeholders play a role.

In the newly emerging context of TDC's, the primary function of evaluation is not accountability. It is an instrument for mutual learning and improving the research effort. It does not mean to strive for a higher ranking in one of the international systems, but it aims at being more effective in addressing a specific societal challenge. As Kuhlmann and Rip (2014) point out, for TDC's or challenge oriented research, new forms and concepts of governance are needed, that go beyond the ideas of the New Public Management. A form of governance that is tentative, dynamic and is a learning process.

Evaluation approaches suited for TDC's are yet underdeveloped, but there are promising examples, such as described by Guba and Lincoln (1989), Patton (1997),

Worthen et al. (1997) and Kuhlmann (2003). One particularly interesting approach comes from the field of development research. Participatory Impact Pathways Analysis (PIPA) is a method that combines evaluation with planning and is intended to mutual learning and looking forward, and involving stakeholders from the start. It focuses on the process and interactions and is instrumental in developing a shared sense of responsibility. It has been developed from earlier ideas in programme theory and pioneered within the Consultative Group on International Agricultural Research (CGIAR) Challenge Programme on Water and Food (Douthwaite et al. 2007a, b). It has been applied in a number of different contexts, primarily to plan and monitor the impact of research for projects in development countries.

The key concept of PIPA is a *theory of change*. A theory of change aims at explaining the logic of activities of a specific programme or project, via a causal narrative. This narrative explains how—via which impact pathway—a specific project or programme is going to make an impact. It includes ideas, expectations and assumptions of the stakeholders involved: how do they envision that their activities will lead to results, and how do these results contribute to impacts? This contrasts the common narrative often used in the wake of the Vannevar Bush report where solutions to societal challenges depend on unfettered basic research. A theory of change opens up this linear narrative and it allows for different contributions coming from different angles in society to participate in the debate about how to achieve a particular desired change. It invites stakeholders to articulate how an impact will be generated, to explain through what steps or outcomes, with whom involved and under what assumptions. It is a way to arrive at knowledge that is socially robust (Nowotny et al. 2001). A theory of change focuses on the ultimate intended impacts and the assumed causal pathways leading towards these impacts. Evaluation or monitoring is focused on evidence about whether or not the impacts have been or are likely to be achieved (Rogers 2014: 10).

PIPA uses logical frameworks for the joint development of a theory of change. A logical framework consists of inputs, activities, outputs, outcomes and impacts of the program or project [see for instance Donovan and Hanney (2011)]. But in order to develop a theory of change, two more steps are necessary. First is the articulation of relations between the elements of a logical framework. Participants discuss the pathways through which they assume that the inputs and activities will lead to outputs and outcomes, or to knowledge and capabilities and the pathways that will result into impact. And second, the articulation of causal assumptions. This relates to the why and how of these impact pathways. How do certain activities lead to specific outputs? Why do certain outcomes lead to a specific impact? These causal relations are the core of the underlying theory of change.

At the start of a project or programme, PIPA uses a participatory workshop to develop a theory of change. During this workshop, researchers and stakeholders jointly develop a theory of change. The participants articulate the ultimate impact, or vision, that they are aiming for, as well as the steps in the process leading towards this impact. A number of questions are central. What outcomes, or changes, are intended? Which actors are necessary to bring about the change? What strategies can be applied, and what project outputs, are needed? How can a specific change be verified? And also: what assumptions that are beyond the control of the project, but that affect the success, should be taken into account? Participants discuss these and other questions in order to develop a joint understanding of the logic of the project or programme. By doing this, they collectively develop a theory of change.

Each participant will bring her or his assumptions, particular logic steps and potential risks for the project or programme to the table. During the workshop, the participants are invited to articulate and share their assumptions; to make their often implicit ideas and



expectations explicit. In other words, the aim of the workshop is to guide the participants from a variety of particular logics, to one shared and articulated logic, or narrative. The joint narrative arguably provides a common impact pathway.

A theory of change and the logical framework are developed in a workshop or series of workshops with representatives of all stakeholders involved in a project, including 'next users' (people and organizations planning to apply the project results), 'end users' (those ultimately benefiting from the research), and other relevant stakeholders (Douthwaite 2009). A workshop is organised at the beginning of a project or programme, and sometimes more workshops follow during the project. The aim is to generate and maintain a shared understanding amongst those involved. This relates to the impacts of the project (what) and to the stakeholders needed in order to realize these impacts, including those that have so far been overlooked or have not yet been included (with whom).

This approach differs in a number of ways from more traditional forms of research evaluation and governance. First, evaluation becomes part of a wider governance process in which the goal is to find common ground between stakeholders. Rather than about accountability towards a funder, it is about mutual learning between interested parties. It is about understanding the implicit logic of a particular research and innovation process and the assumed causal relations between inputs, activities, outputs, outcomes and impact. In case the project or programme develops differently than expected, the theory of change and the logical frame can be discussed and adapted. All stakeholders are involved in this potential change, they share the responsibility for the process and the evaluation of its progress. And with regard to logical frameworks in particular, use differs from other uses in evaluation context, such as in the payback framework. In such cases, the logical framework is developed *ex post* and by an external evaluator. The approach we describe here uses the construction of a logical framework *ex ante*; moreover, the stakeholders themselves construct the logical framework.

A logical framework provides a series of potential indicators to choose from, in each phase of the process. Some of these indicators may relate to the ultimate change that the project or program aims for. Others regard the steps in the process of the project or programme: input, activities, outputs, outcomes and impacts. These (intermediate) indicators are used to monitor the project or programme on the way, *ex durante*. They act as process indicators that can be used to jointly monitor the progress; to jointly understand where the project or program is heading for. Such indicators are to inform stakeholders about the success or failure of the collaboration. This fits a governance model that is tentative, dynamic and aimed at learning. But it differs from the usual practice in academia. They are not meant as stand-alone indicator of research quality, nor are they meant to be used to compare projects or programmes. They have a novel function.

We have used PIPA workshops for a number of related TDC's. From that experience, we have identified a number of challenges, in particular for the academic partners involved.

## 6 Applying PIPA: design and results

One of the authors<sup>1</sup> used the development of a theory of change, based on PIPA, when supporting a regional TDC. This consortium was at the initial stages of a 10 year strategic program. The aim was to strengthen economic and social structures in the region; this would arise from the power of collective knowledge and expertise development. The total

<sup>1</sup> Author Leonie van Drooge with colleague Barend van der Meulen.



budget, including in kind contributions from the various knowledge partners, was over half a billion Euros over a period of 10 years. The intention was to fund a total of 15 different projects within this 10 year program. In every project, relevant institutes of higher education (HEI's), private companies, local governments, NGO's and local associations and foundations were to collaborate. And every project would have to have a relevant societal contribution to the region. For the regional government, the main funder of this program, involvement in a knowledge intensive transdisciplinary program of this scale, was new.

One of the institutes of HEI's involved asked for support to develop a set of indicators for monitoring and evaluation. The HEI attached great importance to the success of the programme and understood that a novel way of monitoring was necessary. It wanted to avoid the situation that only after completion of a project, success, or worse failure, could be determined.

Together with the main stakeholders, the regional government and the main HEI, a number of criteria for indicators were developed. For every project the basic requirements were the same. Indicators:

- function to monitor whether the project will reach the goals set;
- need to enable learning during the project, so as to stimulate changes and improvements;
- should reflect characteristics of the specific project;
- should be realistic to use, and this includes that it should be financially justified to collect evidence;
- need to be endorsed by policy and politics.

We were then asked to support the development of indicators for the first projects to start. We proposed to use the PIPA approach. The joint development of a shared narrative and a monitoring framework, and the use of that framework to learn, were welcomed as very useful aspects. However, it became clear that it was not feasible to conduct workshops of up to 3 days, as described in the PIPA method. The main stakeholders suspected that that there would be insufficient support for such a time investment. A half day workshop was considered to be realistic. Knowing that STEPS (Ely and Oxley 2014) had positive experience with shorter workshops, we agreed to this timeframe.

The first projects to start, were two projects in the field of education and two in the field of medical imaging and materials. In all four projects, knowledge institutes were to collaborate with local organisations. In the two education projects, partners included public organisations, such as schools, private organisations, such as catering companies, as well as a variety of associations (parent-organisations, sports associations, etc.). In the medical imaging and materials projects, partners were predominantly private (companies, as well as business parks).

## 6.1 The workshops

PIPA inspired workshops were organised for each of the four projects. For these, representatives of the main stakeholders were invited. At every workshop, the regional government, the HEI's (principal investigators as well as the board) and other stakeholders were present. The number of attendees, apart from the workshop organisers, was between seven and ten.

Each workshop was organised at the offices of the HEI, at a central location in the capital city of the region. The workshops lasted half a day and they were led by the same two researchers. The workshop started with an introduction of the participants, followed by

an introduction to the method of logical frames. Then the groups were split in two. Attention was paid to the composition of every subgroup: as diverse as possible. Participants were asked to identify the ultimate goal, activities necessary to achieve this, expected results, outcome and impact. Based on these, each subgroup constructed a logical frame. In the final phase, the two subgroups shared their logical frames and developed a common understanding of the pathways and of causalities. The causal relations and the theory of change were discussed. A final round of reflection marked the end of each workshop.

The result of the workshops was a list of goals and sub goals, activities, outputs, outcomes and impacts. This was used to develop a monitoring framework that consisted of criteria and indicators to monitor the progress towards the goals and sub goals identified.

To illustrate the type of goals formulated and the consequential monitoring framework, we present the goals and a number of proposed indicators for two different projects.

## 6.2 Case example: education

One of the education projects is dedicated to the development and application of a new concept for education. The aim is to improve the health of the schoolchildren. The concept affects the common daily routines and schedules of primary schools.

Three sub goals were identified during the workshop: (1) the development of an evidence based concept for this new education paradigm, (2) the sustainable realisation of four schools based on that concept and (3) the consequential application of the concept in other schools, both within the region, as well as beyond the region.

1. The development of the evidence based concept relates to the scientific basis and the effectiveness of the concept, as well as to user costs, ethical issues and legal frameworks. The researchers involved in the field of educational science, focus on the scientific basis and on the effectiveness. To monitor the quality of their research, output indicators such as publications in academic journals, PhD's granted or citations (scientific impact) can be used. It was identified during the workshop, that it was necessary to monitor the progress of the research in other ways as well. Intermediate analyses regarding the effects of the concept were proposed as necessary and relevant activities—and indicators to monitor progress. The validity of the legal framework is a different, yet equally important issue. The rules, regulations and laws regarding education are strict and need to be taken into account. Relevant indicators relate to activities such as the delivery of an inventory of legal and ethical boundaries and possibilities; to outputs such as the design for the management and supervisory board and to outcomes such as the decision on the legal form for such a school.
2. The sustainable realisation of four schools is necessary in order to test and introduce the new concept. It relates to the involvement and commitment of teachers, parents and local organisations. Relevant indicators relate to activities such as organising information meetings, consultations of parents and to outcomes such as the participation of a required minimum percentage of parents.
3. The application of the concept in other schools was not discussed during the workshop due to time constraints. However, it was identified as a goal in the long run.

## 6.3 Case example: new materials

One of the medical projects is dedicated to, again, the development and application of a new paradigm. This relates to novel approaches regarding the production of biomaterials.

The core issue is the *in vivo* production of biomaterials, i.e. by living organisms such as plants. The project involved the erection of an institute centred on this new approach, so as to further develop the novel approach.

The sub-goals were: (1) the development of a sustainable research institute; (2) the achievement of a paradigm change regarding the specific biomaterials and (3) the strengthening of the innovation ecosystem in the region.

1. The development of a sustainable research institute addresses the future, when no extra subsidies are guaranteed. Two aspects were identified as crucial: a solid financial base and ample good staff. Indicators that monitor these aspects relate to inputs, such as total income, number of different contracts, variety of contract partners, the increase in staff and the potential to attract excellent researchers.
2. In order to introduce a novel approach, research and examples are needed and mission work needs to be done. Output indicators such as scientific publications and outcome indicators such as citations can be used to monitor. However, this relates to the quality of the research and to some extent to the acceptance of the new paradigm. Another way to monitor the acceptance of the new paradigm, is through collaborations with scientific peers and industrial stakeholders. Therefore the number of collaborations was defined as an indicator.
3. The establishment of the institute in the region was going to have a positive influence on the regional innovation ecosystem. A series of indicators relating to business activities can be formulated ranging from number of collaborations with regional companies to the relocation of companies towards the region and the foundation of new companies.

## 6.4 General observations

The number of attendees at each of the workshops was between seven and ten. This is very little, given the amount and variety of stakeholders in every project, such as project partners (with whom collaboration throughout the project is necessary), next users (people and organizations that can or should apply the results) and end users (those ultimately benefiting from the research). It became clear that even for the main stakeholders involved, it was difficult to convincingly invite core stakeholders to the workshop.

During the workshops and through the discussions, it became clear that the responsibility for each of the projects is shared with many different stakeholders. They have different interests and different ideas about quality and relevance. Representatives of the regional government explained in a convincing way that next to good research it was imperative to have a convincing narrative for the regional politicians. They stressed that the politicians are interested in a more detailed narrative than the Bush narrative we referred to above.

Also, two very different logics became apparent. For the local government, the logical frame aims at how the programme as a whole and how each of the individual projects contribute to strengthening the economic and social structures in the region. For the researchers, the logical frame is about how to get funding that supports the kind of research that can lead to societal applications and to entrepreneurial activity.

An imbalance between the stakeholders came to the fore, regarding the robustness of the quality control. For example, when discussing the quality of research, the non-scientists found it difficult to identify ways to assess the quality and relevance of research in relation to the project, and thus to formulate indicators. The participating scientists rather easily

formulated quality indicators (peer reviewed publications, PhD granted, conferences organised). This establishes research quality but not relevance to the project. When non-scientific aspects were discussed, such as the reallocation of certain companies to a campus or the willingness of a school to participate, the societal partners were able to identify milestones to assess the progress towards the ultimate goal.

It became clear from the workshop that it takes time to understand and accept each other's approaches to quality and relevance control. In this case, it was the first time that the project aims and the impact narrative were discussed on such an intensive scale, and with an extended group of participants. The workshop proved to be useful as a means to bring stakeholders together and discuss the aims of the funding and the project. But it became clear as well that it takes more time in order to develop a true shared understanding and a shared responsibility.

## 7 Discussion and conclusion

Conducting research in a transdisciplinary context is challenging. A TDC incorporates a variety of societal stakeholders, each with their own assumptions, ideas, goals and expectations. They all bring in different expertise that somehow needs to be attuned. This has consequences for the governance of TDC's, including evaluation. To do justice to this joint process aimed at societal impact, evaluation can only be a joint effort too. Such evaluation is aimed in the first place at improving the collaborative understanding of the joint process and secondly at the progress towards the common societal goal. The latter means a focus on small steps that stakeholders make towards the common goal. These may vary from a scientific article to the setup of a joint facility or testing ground to public engagement. Scientific research is only a part of the process, and evaluation includes more aspects than scientific excellence and scientific impact. In order to assess complex endeavours as TDC's, changes are required to the current practice of research evaluation. Evaluation needs to be more comprehensive and part of a joint governance process of a TDC.

PIPA is a method that suit TDC's. The participants or stakeholders (including the researchers, funders and users) involved in a TDC collaborate and gather from the start. A kick off workshop is organized to jointly discuss the societal change they want to achieve. As we have learned, it is not obvious for stakeholders, including core partners, to spend a substantive amount of time together to discuss a major project. Major in terms of time, finance and ambitions. The workshops we organized were the first instances that the core stakeholders present met at the same time. In all four workshops however, the stakeholders responded in a positive way to the joint discussions of the intentions, goals and process of the project.

In the workshop, the stakeholders construct theory of change. This is a narrative about the intended impact and goals, and about the assumptions, and processes that will lead towards that impact. Stakeholders next need to agree about the intermediate steps to be taken to reach that goal. In PIPA, this is formulated in terms of a logical framework. We experienced that stakeholders have different expectations regarding the organisation of the project, different ideas regarding the impact and use different logics to the problem at hand. The articulation of these differences was experienced by participants as a useful insight. The need for mutual learning and understanding became clear. But mutual learning

and understanding takes time. Time that stakeholders, in particular researchers, are not used to spend to such activities.

When stakeholders agree on the theory of change and the logical frame, criteria and indicators can be formulated that fit the shared goals. What we have learned is that the theory of change provides a good and systematic narrative. This narrative helps a lot, especially when supported by concrete evidence in the form of quantitative or qualitative indicators.

The indicators differ in function from what one is used to: they focus on the process of collaboration, and look forward and not back. They aim at monitoring and assessing the steps in the process towards the final goal. Their main goal is not to measure scientific excellence, but to understand whether societal change is achieved. Excellent scientific results are needed, though, to reach that societal goal, but it is not the main objective.

Furthermore, a range of indicators can be relevant, and somehow these have to be attuned in the assessment. TDC's are complex endeavours that entail different types of knowledge and expertise; they aim at a mix of scientific and societal goals. In both cases presented above, it became clear that for most stakeholders the goals of the projects related predominately to societal goals. As a consequence, the traditional scientific indicators were not very useful, while other indicators used (the percentage of parents willing to participate in the education case, or the number of companies that relocate to the region in the materials case) did not relate to scientific practices. It is a matter of governance how to deal with these differences in an acceptable way. Again, good narratives, preferably supported by robust evidence, may help in these cases.

Coming back to the crossroads of science and society that HEI's are operating on. In order to change the culture and make it fit for TDC's, changes are needed at all levels of the science system, as for example Kuhlmann and Rip (2014) and Schot and Steinmuller (2016) have pointed out. The experience with PIPA emphasizes this need for change. From the four workshops we learn how large the gap is between everyday scientific practice and the needs of TDC's. Whether in future PIPA will be used in evaluation of TDC's, or other methods will be developed, changes need to include the function of evaluation—mutual learning instead of accountability—and the use of narratives supported by novel qualitative or quantitative indicators. This requires a major change by scientists and funders: the abandonment of the linear Bush model of science. And although in the literature on science, technology and innovation studies, this model has been diagnosed as far beyond expiration date, the experience we describe in this article has learnt that it still provides a strong narrative for scientists.

**Acknowledgements** We thank two anonymous reviewers for their comments; Barend van der Meulen with whom the pilot for the regional programme was conducted; all participants in the regional programme; and, for their substantive contribution to the section on the changing policy context, Jasper Deuten, Boru Douthwaite, Adrian Ely, Patricia Faasse, Laurens Hessels and Jordi Molas-Gallart (in alphabetical order). An earlier version of this paper has been presented at the conference *Impacts of agricultural research – towards an approach of societal values*, organised by INRA, Paris, November 2015.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Benedictus, R., & Miedema, F. (2016). Redefine excellence: Fix incentives to fix science. *Nature*, 538, 453–455.
- Blankesteyn, M., Munnichs, G. M., & van Drooge, L. (2014). *Contested science: Public controversies about science and policy*. The Hague: Rathenau Instituut.
- Blume, S. S., & Spaapen, J. B. (1988). External assessment and “conditional financing” of research in Dutch universities. *Minerva*, 26(1), 1–30.
- Bush, V. (1945). *Science the endless Frontier. A report to the president*. Washington, DC: United States Government Printing Office.
- Clark, W. (2006). *Academic Charisma and the origin of the research university*. Chicago: University of Chicago Press.
- de Jong, S., Smit, J., & van Drooge, L. (2016). Scientists’ response to societal impact policies: A policy paradox. *Science and Public Policy*, 43, 102–114.
- Donovan, C., & Hanney, S. (2011). The ‘payback framework’ explained. *Research Evaluation*, 20(3), 181–183.
- Douthwaite, B., Alvarez, B. S., Cook, S., Davies, R., George, P., Howell, J., et al. (2007a). Participatory impact pathways analysis: A practical application of program theory in research-for-development. *Canadian Journal of Program Evaluation*, 22(2), 127–159.
- Douthwaite, B., Alvarez, S., Keatinge, J., Mackay, R., Thiele, G., & Watts, J. (2009). Participatory impact pathways analysis (PIPA) and research priority assessment. In D. A. Raitzer & G. W. Norton (Eds.), *Prioritising agricultural research for development* (pp. 8–24). New York: CABI.
- Douthwaite, B., Schulz, S., Adetunji, S. O., & Ellis-Jones, J. (2007b). Impact pathway evaluation of an integrated *Striga hermonthica* control project in Northern Nigeria. *Agricultural Systems*, 92, 201–222.
- Edler, J., & Kuhlmann, S. (2008). Coordination within fragmentation: Governance in knowledge policy in the German federal system. *Science and Public Policy*, 35(4), 265–276.
- Ely, A. Y., & Oxley, N. (2014) STEPS centre research: Our approach to impact. *STEPS Working Paper* 60. Brighton: STEPS Centre.
- European Commission (2017a) Societal challenges. <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges>. Accessed 3 April 2017.
- European Commission (2017b) Joint programming. [http://ec.europa.eu/research/era/joint-programming\\_en.html](http://ec.europa.eu/research/era/joint-programming_en.html) Accessed 2 April 2017.
- Expert Group on Policy Indicators for RRI. (2015). *Indicators for promoting and monitoring responsible research and innovation: Report from the expert group on policy indicators for RRI (EUR 26866 EN)*. Brussels: European Commission, DG R&I.
- Freeman, C., & Soete, L. (1997). *The economics of industrial innovation* (3rd ed.). Cambridge, MA: MIT Press.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.
- Gray, D. O. (2011). Cross-sector research collaboration in the USA: A national innovation systems perspective. *Science and Public Policy*, 38(2), 123–133.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park: Sage.
- Hemlin, S., & Rasmussen, S. (2006). The shift in academic quality control. *Science, Technology and Human Values*, 31(2), 173–198.
- Hessels, L. K. (2010). *Science and the struggle for relevance Ph.D. thesis*. Utrecht: Utrecht University.
- Hessels, L. K., van Lente, H., & Smits, R. (2009). In search of relevance: The changing contract between science and society. *Science and Public Policy*, 36(5), 387–401.
- Hoppe, R. (2010). *The governance of problems—puzzling, powering and participation*. Bristol: Voss.
- Joly, P. B., Matt, M., Gaunand, A., Colinet, L., Larédo, P., & Lemarié, S. (2015). ASIRPA: A comprehensive theory-based approach to assess societal impacts of a research organization. *Research Evaluation*. doi:10.1093/reseval/rvv015.
- Kloet, R. R., et al. (2013). Understanding constraints in the dynamics of a research program intended as a niche innovation. *Science and Public Policy*, 40(2), 206–218.
- Kuhlmann, S. (2003). Evaluation as a source of “strategic intelligence”. In Ph Shapira & S. Kuhlmann (Eds.), *Learning from science and technology policy evaluation: Experiences from the United States and Europe*. Cheltenham: E. Elgar Publishing.
- Kuhlmann, S., & Rip, A. (2014). The challenge of addressing grand challenges: A think piece on how innovation can be driven towards the “Grand Challenges” as defined under the prospective European Union Framework Programme Horizon 2020. [http://ec.europa.eu/research/innovation-union/pdf/expert-groups/The\\_challenge\\_of\\_addressing\\_Grand\\_Challenges.pdf](http://ec.europa.eu/research/innovation-union/pdf/expert-groups/The_challenge_of_addressing_Grand_Challenges.pdf).

- Lam, A. (2011). What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'? *Research Policy*, 40, 1354–1368.
- Nelson, R. (1977). *The moon and the ghetto: An essay on policy analysis*. New York: W. W. Norton.
- Nowotny, H., Scott, P., & Gibbons, M. T. (2001). *Re-thinking science: Knowledge and the public in an age of uncertainty*. London: Wiley.
- OECD. (2012). *Meeting global challenges through better governance: International co-operation in science, technology and innovation*. Paris: OECD Publishing.
- Patton, M. Q. (1997). *Utilization-focused evaluation. The new century text*. Thousand Oaks: Sage.
- Pielke, R. A. (2010). *The climate fix: What scientists and politicians won't tell you about global warming*. New York: Basic Books.
- Rip, A. (2004). Strategic research, post-modern universities and research training. *Higher Education Policy*, 17(2), 153–166.
- Rogers, P. (2014). *Theory of change. Methodological briefs: Impact evaluation 2*. Florence: UNICEF Office of Research.
- Sarewitz, D. (2016). Saving science. *The New Atlantis*, 49, 5–40.
- Schot, J., & Steinmueller, W. E. (2016). *Framing innovation policy for transformative change: Innovation policy 3.0*. Brighton: University of Sussex.
- Scriven, M. (1991). *Evaluation thesaurus* (Vol. 4). Newbury Park, CA: Sage.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation*, 2(4), 393–404.
- Spaapen, J. B., & van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. *Research Evaluation*, 20(3), 211–218.
- Stirling, A. (2006). *From science and society to science in society: Towards a framework for "co-operative research"*. Luxemburg: Office for Official Publications of the European Communities.
- van den Akker, W., & Spaapen, J. B. (2017). Productive interactions: Societal impact of academic research in the knowledge society. *LERU Position Paper*. Leuven: LERU.
- van der Veen, G. (2005). *Evaluation leading technological institutes: Final report*. Amsterdam: Technopolis Group.
- Whitley, R., & Gläser, J. (Eds.). (2007). *The changing governance of the sciences*. Dordrecht: Springer.
- Wissema, J. G. (2009). *Towards the third generation university: Managing the university in transition*. Cheltenham: Edward Elgar.
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation. Alternative approaches and practical guidelines* (2nd ed.). White Plains, NY: Longman.