



Evaluating co-creation of knowledge: from quality criteria and indicators to methods

Susanne Schuck-Zöllner, Jörg Cortekar, and Daniela Jacob

Climate Service Center Germany, Helmholtz-Zentrum Geesthacht, 20095 Hamburg, Germany

Correspondence to: Susanne Schuck-Zöllner (susanne.schuck@hzg.de)

Received: 21 December 2016 – Revised: 23 August 2017 – Accepted: 7 September 2017 – Published: 3 November 2017

Abstract. Basic research in the natural sciences rests on a long tradition of evaluation. However, since the San Francisco Declaration on Research Assessment (DORA) came out in 2012, there has been intense discussion in the natural sciences, above all amongst researchers and funding agencies in the different fields of applied research and scientific service. This discussion was intensified when climate services and other fields, used to make users participate in research and development activities (co-creation), demanded new evaluation methods appropriate to this new research mode.

This paper starts by describing a comprehensive and interdisciplinary literature overview of indicators to evaluate co-creation of knowledge, including the different fields of integrated knowledge production. Then the authors harmonize the different elements of evaluation from literature in an evaluation “cascade” that scales down from very general evaluation dimensions to tangible assessment methods. They describe evaluation indicators already being documented and include a mixture of different assessment methods for two exemplary criteria. It is shown what can be deduced from already existing methodology for climate services and envisaged how climate services can further to develop their specific evaluation method.

1 Introduction

The production of climate-related information, as well as the assessment of possible impacts and adaptation options or even scenarios and strategies, depends on the participation of practice partners (Schuck-Zöllner et al., 2014). Their involvement in research and development processes helps to ensure the usability of results. Corresponding to these experiences, the European commission states that climate services are by nature inter- and transdisciplinary (European Commission, 2015, p. 22). The core idea of transdisciplinarity was intensely discussed in 2000 at a congress in Zurich, Switzerland, pledging academic disciplines to work together with practitioners to answer real-world problems. Jahn et al. (2012, p. 4) give a comprehensive definition: “Transdisciplinarity is a reflexive research approach that addresses societal problems by means of interdisciplinary collaboration as well as the collaboration between researchers and extra-scientific actors; its aim is to enable mutual learning processes between science and society; integration is the main cognitive challenge of the research process.” Launched

by the Future Earth community (www.futureearth.org), the terms “co-design” and “co-production of knowledge” are spreading as well (Mauser et al., 2013). Scientists and stakeholders from practice or politics continuously work together: “The process of co-creation of knowledge consists of three fundamental steps throughout which both academia and stakeholders are involved to varying degrees: co-design, co-production and co-dissemination” (Mauser et al., 2013, pp. 427, 428). Wall et al. (2017) gave a comprehensive overview of the different concepts and terms in the area of transdisciplinary research worldwide.

The transdisciplinary research mode and co-creation of knowledge as its methodological approach aims to solve societal problems. Basic research, however, is focused primarily on the scientific impact. The number of publications, the ranking of the respective journals and the number of quotations are used to evaluate the impact onto the scientific world. Therefore, basic research follows different guiding principles and goals than transdisciplinary approaches and has to be evaluated in another way. “Multiple forms of co-operation,

differentiation and integration, methods and theories are significant for such [transdisciplinary] projects. So conventional methods of disciplinary evaluation cannot be transferred and applied directly” (Bergmann et al., 2005, p. 7). Jahn and Keil (2015) state that “there is still no generally accepted quality standard” in regards to the requirements of socially responsible research and are afraid that the missing incentives for researchers in this field “obstruct the proliferation of transdisciplinarity” (p. 196).

This has been discussed in literature since the beginning of the 21st century (Klein, 2008; Kaufmann and Kasztler, 2009) in sustainability science, health care and in overarching scientific fields like scientometrics, sociology of science and innovation studies. Rafols et al. (2012) showed how the bibliometric evaluation method might panelize inter- and transdisciplinary approaches. Meanwhile the need for new evaluation criteria and methodologies is absolutely agreed upon (Wolf et al., 2013; Jahn and Keil, 2015; Wall et al., 2017).

The high percentage of co-creation in climate services encouraged the authors of this paper to base their discussion on transdisciplinarity and co-creation, their quality and evaluation to draw conclusions for the evaluation of the fairly new field of climate services. They aim to bring together the different levels of evaluation elements found in literature and present the idea of an evaluation cascade. This paper does not attempt to find a complete set of criteria but rather shows how to apply the resulting cascade in practice. A literature overview is our first approach to the challenge and delivers already existing quality dimensions, criteria and indicators to those domains in which co-creation of knowledge is applied. The different approaches in the current discussion show the elements for a workflow to make evaluation easier to handle: the evaluation cascade. In line with this, the authors demonstrate in a second step how two exemplary criteria could be assessed in practice. They give a potential description of how the evaluation cascade could be applied to two climate service products and propose appropriate assessment methods. In the end the authors suggest how to proceed towards a framework for the evaluation of co-creation products and projects.

2 Literature overview

2.1 Key questions

In 2015, the Climate Service Center Germany (GERICS) undertook a literature survey, aiming to identify existing evaluation indicators and assessment methods. As there is a much longer history of working with practice partners in other research fields (Bergmann et al., 2012), such as public health and sustainability sciences, all scientific literature was integrated. The long-term objective was to find out and critically reflect upon which systems should be established in order to evaluate co-creation and transfer this to the whole field of climate services.

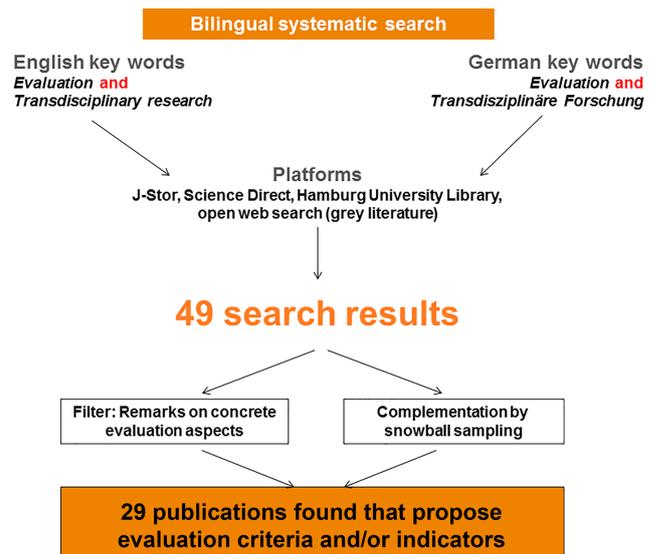


Figure 1. Search method for literature overview in the field of transdisciplinary research.

2.2 Method

As a first step an overview of the state-of-the-art discussion was needed. Thus a literature collection was started in 2015, covering the past decade (2005–2015). The search was done without any restrictions on thematic fields. Two key items (“evaluation” and “transdisciplinary research”) were decided on, in both English and German, and the search was reduced to combinations thereof (Fig. 1). Titles, headlines and full texts of publications were covered. The 49 results originated from very different fields:

- sociology of science and epistemology (17 articles)
- sustainability and ecology (17 articles)
- health (9 articles)
- technology (3 articles)
- urban studies (3 articles).

The result shows that the search was general enough to cover different disciplines and the neighbouring fields. As the mode of transdisciplinary research and its discussion is anchored in the social–ecological field, it is logical that most of the articles originate there – except for the overarching epistemological area. Regarding the unique and relatively new field of climate services, no result could be found by this literature search.

The objective of the literature overview was not to cover the whole discussion in terms of scientometrics, which is indeed very broad, but rather to concentrate on the evaluation of co-creation in detail. Scanning the 49 articles referring to concrete evaluation aspects, a selection of the relevant publications was made. The pool was further complemented with

publications that were already known or that showed up during work (snowball sampling). The compilation ended up with 29 publications that propose concrete evaluation aspects on the levels that this paper calls “criterion” and “indicator” (Climate Service Center Germany, 2017).

2.3 Results

Evaluation can firstly be differentiated alongside its object and the literature overview covered the objects of evaluation. Among others either the process of co-creational research or its final product can be evaluated (Fig. 2, right-hand and left-hand side, respectively). Secondly, there are different phases of research projects during which evaluation can take place: *ex ante*, intermediate or formative, and *ex post* evaluation.

In general, the authors found that the different application fields of transdisciplinary research follow very similar ideas in the choice of evaluation criteria and indicators. So the assumption was confirmed that climate services might benefit from other research fields in this respect. The assessment of impacts on a societal scale, however, is barely discussed in terms of indicators. Except for very few contributions (i.e. Godin and Doré, 2005; Spaapen et al., 2007; Walter et al., 2007) the publications do not offer suggestions of assessment criteria of long-term effects inside or outside the scientific area. Therefore, the authors decided to concentrate on the evaluation of the research process itself (Fig. 2: “Dealing with the problem”) and on the evaluation of outcome (short- and medium-term effects).

The articles of our overview show very different approaches – not surprising considering the different research communities given. Some authors, for example, concentrate on good quality (Schuck-Zöllner et al., 2018) and best-practice (like Stauffacher et al., 2008). Others describe processes of formative evaluation (e.g. Bergmann et al., 2005) or aim for guidelines for project managers (e.g. Jahn and Keil, 2015). Still others draft evaluation frameworks (Vaughan and Desai, 2014). There were almost no cross references among the articles (except for Wolf et al., 2013, who gave an overview of the different evaluation concepts). Therefore, the authors wanted to bring together those different approaches and suggest a common terminology. The possibility of an overarching evaluation scheme is shown. The paper focuses on *ex post* evaluation. Based on literature analysis, the authors identified different factors of evaluation.

As the literature overview was only the first step and not the core issue of this article, the authors went beyond the articles they had found, scanning literature for the overview mentioned above.

3 Evaluation cascade

Some of the articles propose a hierarchic structure to scale down from the very general towards more and more detailed scales. Different names and terms are used for the

most general level, though the ideas seem similar. Bergmann et al. (2005), for example, only differentiate between “basic criteria” instead of “dimensions” and “detailed criteria”. Klein (2008) calls this level “principles”, while Walter et al. (2007) uses “impact categories”. The authors follow Jahn and Keil (2015) and Hassenforder et al. (2015), who use the term “dimension”, as it seems quite suitable for this general level. Jahn and Keil (2015) identify further nine different dimensions to evaluate transdisciplinary research.

Accordingly, the quality of the research problem can be expressed in three dimensions:

- systemic quality
- scale spanning quality
- prospective quality.

The quality of the research process composes of three dimensions:

- context-specific quality
- integrative quality
- method-based quality.

The quality of the research results can be looked upon in three further dimensions, as well:

- critical-reflexive quality
- normative quality
- impact-oriented quality.

The authors refer to these nine dimensions in this paper. In other articles this general level is subcategorized slightly differently, more so as they follow different aspects (e.g. Bergmann et al., 2005, concentrate on formative evaluation; Godin and Doré, 2005, only look at the societal impact; Hassenforder et al., 2015, focus on participatory processes). To make the nine dimensions mentioned above applicable they have to be divided into smaller subcategories. For this, the authors designed a scale, the evaluation cascade, that ranges from the very general aspect of dimensions to indicators and assessment methods (see Fig. 3).

In an attempt to demonstrate the potential of the theoretical idea of a cascade in real life, the different evaluation aspects from literature were collected and filled in an evaluation scheme alongside the evaluation cascade. The level underneath the more general category “dimension” gives an indication of which specific aspect should be assessed. The authors found the term “criterion” easy to understand and widely used (i.e. Bergmann et al., 2005; Klein, 2008; Wolf et al., 2013), as well as “indicator” (for example Klein (2008) in contrast to Masse et al. (2008), who still use the very general “item”). Indicators in this sense refer to a unit to assess a specific state of something. Wall et al. (2017) mix up what

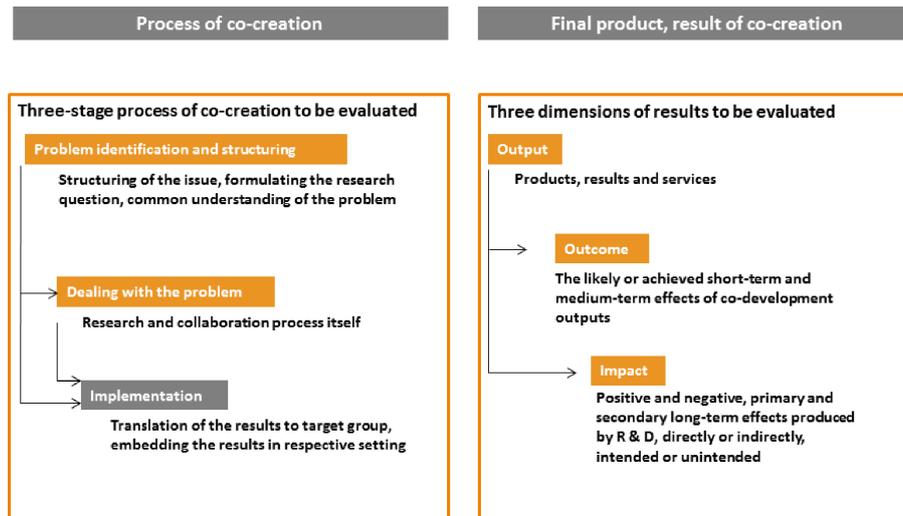


Figure 2. The process of co-creation (on the left) spreads up into three consecutive phases (Pohl and Hirsch Hadorn, 2007), each of them can be evaluated. In terms of product or result of this process, either the “output”, or “outcome” or the “impact” (OECD, 2002) might be assessed. Describing texts adapted to the issue by the authors.

this text calls criteria and indicators and only differentiate between “components” (similar level like the suggested “dimensions”) and indicators.

Which criteria and indicators can be found in literature? Different kinds of criteria are scattered about in many publications (except for Bergmann et al., 2005). Nevertheless, in most of the papers very few indicators could be found, and concrete measurement methods, which might finish the cascade on the micro scale, are missing. If questions to assess the indicators are posed, they are qualitative and hardly measurable (e.g. Bergmann, 2005; Jahn and Keil, 2015).

In any case, the evaluation concepts identified in literature are dedicated to different kinds of evaluation. Some refer to ex post evaluation (i.e. Wolf et al., 2013). The contributions of Bergmann et al. (2005) and Jahn and Keil (2015), for example, are dedicated to formative evaluation. The concepts follow very different objectives and the resulting criteria are very diverse (Wolf et al., 2013). The few publications dealing with criteria in a concrete manner show an enormous pool (i.e. Bergmann et al., 2005; Wall et al., 2017). Nevertheless, it seems important to compile all of the ideas in one overall representation.

On the basis of the evaluation cascade, several tables were designed (i.e. Fig. 4) and filled in as far as possible. The column “indicators”, however, cannot be filled at all on the basis of literature overview, because many papers leave out the level of indicators completely (e.g. Bergmann et al., 2005; Jahn and Keil, 2015). Therefore, the entries here and in the “methods” column on the right-hand side are a proposition of the authors, resulting from experiences in the development of climate service products.

Two possibilities of measuring single quality criteria are demonstrated in the following – one assessing the research



Figure 3. The evaluation cascade from the general scale to assessment methods in detail; whereas dimensions, criteria and indicators are covered by literature – although with very different intensity and to a very different extent – the methods mentioned in this article are suggestions by the author (in a separate column in grey).

and development process and the other assessing their results. Therefore, the authors can only show some cutouts to illustrate a way of applying the evaluation cascade and propose the design of a possible evaluation scheme. For this purpose they go beyond the literature survey and add ideas of assessment methods to the state-of-the-art discussion.

3.1 Evaluation of processes

All three steps of the co-creation process (Fig. 2, left-hand side) might be subject to evaluation. This paper leaves out the problem identification and structuring phase and turns directly to the core research process: dealing with the problem. The articles of our sample provide three different dimensions to assess the transdisciplinary research and development process according to the list of Jahn and Keil (2015) (see Fig. 4) and several criteria. Those dimensions are fully described in various papers (i.e. Bergmann et al., 2012; Hassenforder et al., 2015) and mirror well the state-of-the-art discussion in 2015.

In the following, the elements of the cascade are demonstrated for one single criterion belonging to the dimension

Dimension	Criterion (examples)	Indicator (examples)	Method (examples)
Context specific quality	<ul style="list-style-type: none"> Relation to concrete problems Space for reflection and self-assessment Handling of cognitive boundaries 		
Integrative quality	<ul style="list-style-type: none"> Setting the scene for co-production of knowledge Recognition of all kinds of knowledge and different normative systems Transparency of mutual expectations Transparency of different roles 	<ul style="list-style-type: none"> Intensity and duration of involvement 	<ul style="list-style-type: none"> Interviews of project participants
Method-based quality	<ul style="list-style-type: none"> Accountable, transparent generation and evaluation of knowledge Professional planning and management Achievement of joint problem ownership Organizational support 		

Figure 4. The example demonstrates how one criterion out of the dimension “integrative quality” within the research and development process might be assessed. It only shows a selection of possible criteria. The interviews might use the Likert scale (1–5) to quantify the result in some way.

of “integrative process quality” (Masse et al., 2008). “Setting the scene” is a quite important criterion in terms of co-creation of knowledge (Schuck-Zöllner et al., 2018), because it is key for successful involvement of practitioners. Walter et al. (2007) could show that the indicator “involvement” influences the outcome of co-creation positively.

To figure out how the scene was set, interviews with the co-creation participants would be an appropriate method. A Likert scale (1 to 5) could be used to make the answers measurable. The survey could ask questions to the participants of practice such as

- How content are you with the communication set-up on the whole?
- Was enough time provided for involvement?
- Was the material shared with all participants appropriate and the language understandable?
- Was a partnership on equal footing promoted?

Questions to the co-creation participants from science might be

- How content are you with the communication set-up on the whole?
- How content are you with the moderation of the involvement process?
- Did you feel understood in terms of scientific rules and limitations?

Answers to a set of questions like these allow for assessing the quality of the scene for co-creation. However, time and resources are needed for a survey like this.

Dimension	Criterion (examples)	Indicator (examples)	Method (examples)
Quality of output	<ul style="list-style-type: none"> Productivity Elaboration Innovation Critical reflection Safeguarding knowledge and transparent documentation 	<ul style="list-style-type: none"> Number of reviewed papers Number of conference contributions Grey literature Number of web accesses Comprehensibility Advances in knowledge Availability of detailed final report and document server 	<ul style="list-style-type: none"> Check documentation Check documentation Check documentation Web statistics
Quality of outcome	<ul style="list-style-type: none"> Development of new products and services Reception of results, scientific acknowledgement 	<ul style="list-style-type: none"> Usability Citations 	<ul style="list-style-type: none"> Survey Bibliometrics

Figure 5. The example shows how the quality of outcome could be assessed by the indicator “usability” of new products and services.

3.1.1 Example 1: Development of adaptation options with municipal administration

Two potential examples illustrate how the downscaling from dimension to assessment methods via the evaluation cascade might work. To assess how satisfied the scientists on the one hand and the practitioners from administration on the other hand are with the scene for co-creation of knowledge, surveys or interviews could be conducted, including questions like those mentioned above. Some more criteria with respect to indicators and assessment tools would have to be added to assess the integrative quality, e.g. whether there was transparency about the different roles and expectations of participants. In this case, a formative evaluation could also help the project managers to rearrange parts of the integration concept.

3.2 Evaluation of results

Assessing the results of a research and development process, there are three different dimensions to name (for definitions see OECD, 2002): (1) quality of output, (2) quality of outcome, and (3) impact. For reasons mentioned previously, the authors only consider the quality of outcome. Whereas many indicators for the quality of output are measurable (e.g. the number of users might prove good quality of content), the quality of outcome is less tangible. One of the most important indicators is, of course, the usability of new products and services (Fig. 5). Among other methods, a user survey might help to judge usability.

To give an idea, some questions to the users are mentioned here as examples. The users might choose among three answers and tick boxes to respond (a) yes, (b) sometimes or partly, or (c) no:

- Do you use the new product X?

- Does it facilitate your work?
- Is it easy to handle?
- Does the product allow for new features or findings?

To gain a more complex view some qualitative questions should not be missing, such as

- What is the overall benefit, using the product?

The core question could be answered on a Likert scale:

- How do you judge the quality of the product on the whole?

Using the evaluation cascade in the manner described here makes single quality dimensions easier to assess.

3.2.1 Example 2: Development of an interactive web portal

An interactive portal like the Impact2C weatlas (<https://www.atlas.impact2c.eu/en/>) could be evaluated by sending a questionnaire to its users, containing the questions above. Otherwise, an online survey could be implemented on the platform. In this case, the target group of the survey was not the developers like in example 1, but the users or potential users. Depending on the aim of the survey, other questions could be added.

3.3 Summary

Since there is no one-size-fits-all scheme, every project or product needs its own set of evaluation criteria (Jahn and Keil, 2015). At best, this set can be negotiated with the project leaders or even already be suggested in the process of designing the project (Daschkeit and Loibl, 2007).

It is undeniable that both the search for and definition of appropriate indicators and measurements for transdisciplinary research take a long time and require intense discussion.

Whereas in literature qualitative interviews are often pled for (Bergmann et al., 2005; Jahn and Keil, 2015), this seems unsatisfying for an ex post evaluation, since all those interviews mirror the perception of the interviewees more than they can serve as an objective evaluation result. Therefore, a mixture of assessment methods was addressed: both qualitative and quantitative. The authors suggest Likert scales as a method, which would help to quantify qualitative interviews more precisely instead of the often proposed “yes” and “no” answers (Jahn and Keil, 2015).

Two criteria could be scaled down to assessment methods and two examples of products were mentioned. The different levels of evaluation were brought together in the cascade and examples were given on how to apply it rather than making the entire range of criteria complete.

Of course, those two criteria cannot satisfyingly serve to evaluate a product or project on the whole. A broad set of criteria has to be chosen, representing the dimensions that refer to the aim of the respective evaluation object. To show an example case of a product completely being played out in evaluation is not within the scope of this article.

The tables generated in this paper cannot serve as a one-size-fits-all scheme. In general for every evaluation process an individual set of indicators has to be chosen specifically for this very cause. By creating this set, both the guiding principles of transdisciplinary research and the goals of the specific project or product are the core.

4 Conclusions and outlook

Literature provides a remarkable range of quality dimensions and evaluation criteria referring to the process and result of research and development activities. Still, in the articles societal impact in the long run is not yet an issue.

Following literature, an evaluation cascade can be developed that scales evaluation down from quality dimensions to quality indicators and assessment methods. The aspects were collected over many different research fields. In spite of nearly no cross references between the different approaches the basic ideas were similar enough to be integrated. This experience reveals the necessity of looking upon transdisciplinarity and its evaluation as a new, overarching boundary methodology.

The evaluation cascade was exemplarily described for two examples to demonstrate the possible evaluation of one single criterion each, proposing a mixed methodology. Interviews of project participants or users will be part of most evaluations. These surveys might take 2 to 3 months. This makes clear that evaluation of co-creation processes and their results is very time-consuming and extensive. What is more, evaluation of the medium-term outcome and societal impact only makes sense a few months or even years after finishing the project or product. Therefore, funding institutions should be aware of this and provide enough time for evaluation or allow for a subsequent evaluation phase.

The evaluation cascade might help scientists from all fields of socially responsible research reflect on possible ex post assessment activities, as well as prepare for formative evaluation during the project. Until now product development processes or even products have not often been assessed in a scientific way. To ensure objectivity we recommend including third parties. Above all, it can serve as a scheme to compile the different evaluation criteria and assessment methods from literature. Such a compilation over the whole range of the evaluation cascade would help further discussion and might lead to an overall – in the sense of overarching the different research fields – operational evaluation framework: a challenge for the whole community that applies co-creation methods. This framework might provide a broad pallet of cri-

teria and indicators that help to combine the appropriate elements for the respective evaluation measure. The emerging field of climate services can thus enormously benefit from the work already done in the other fields, above all sustainability and public health. Real case examples of climate services evaluation might be developed to further approve the different dimensions and criteria and to ensure quality.

A quite innovative concept has been implemented in the Netherlands to harmonize the evaluation of research on the national level. This was not covered by our literature review, but it indeed takes transdisciplinary approaches into account (Spaapen et al., 2007). Spaapen et al. (2007) “distinguish a number of social domains in which researchers operate” (p. 7). Alongside these domains they define their dimensions and scale down to criteria and assessment. The most challenging aspect of this system is that they try to combine different criteria to make evaluation more easy to handle. After some years of application it would be important to take a second look at Dutch experiences. In the UK, another pioneer country in the evaluation of social responsible research, special guidelines were published to evaluate public engagement activities in research (Research Council UK, 2011). It seems a valuable task to dive deeper into those two international evaluation schemes.

How far practitioners should be involved in the concept of evaluation processes is still not expressively discussed. Should they only take part in the evaluation surveys or rather become part of the evaluation committees? Finally, should practitioners be involved not only in the evaluation of single projects but also in the further development of an overarching evaluation scheme? This would be a task for a new approach of “co-evaluation”. However, it is clear that assessing co-creation processes and its products will still keep the community busy for awhile.

Data availability. No data sets were used in this article.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue “16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC)”. It is a result of the 16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), Trieste, Italy, 12–16 September 2016.

Acknowledgements. The authors want to thank Stefan Füsers for his preparatory work.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

Edited by: Klaus Grosfeld

Reviewed by: Vera Rabelt and one anonymous referee

References

- Bergmann, M., Jahn, T., Knobloch, T., Krohn, W., Pohl, C., Brohmann, B., Hoffmann, E., Loibl, M. C., Rehaag, R., Schramm, E., and Voß, J. P.: Quality Criteria of Transdisciplinary Research – A Guide for the Formative Evaluation of Research Projects, ISOE-Studientexte, 13, Institute for Social-Ecological Research (ISOE), 78 pp., Frankfurt/M., Germany, 2005.
- Bergmann, M., Jahn, T., Knobloch, T., Krohn, W., Pohl, C., and Schramm, E.: Methods for Transdisciplinary Research, Campus, Frankfurt/M., Germany, 2012.
- Climate Service Center Germany: 29 publications referring to the evaluation of criteria and indicators, measuring co-creation processes, available at: http://www.gerics.de/imperia/md/content/csc/gerics/literaturliste_review.pdf, last access: 26 October 2017.
- Daschkeit, A. and Loibl, M. C.: Projektspezifische Evaluation human-ökologischer Nachhaltigkeitsforschung, in: Evaluation inter- und transdisziplinärer Forschung, edited by: Pohl, C. and Stoll-Kleemann, S., Edition Humanökologie, 5, Oekom, München, Germany, 157–173, 2007.
- European Commission: A European research and innovation Roadmap for Climate Services, Bruxelles, available at: <http://www.wmo.int> (last access: 20 December 2016), 2015.
- Godin, B. and Doré, C.: Measuring the impacts of science – beyond the economic dimension, Urbanisation INRS, Culture et Société, available at: http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf (last access: 20 December 2016), 2005.
- Hassenforder, E., Smajgl, A., and Ward, J.: Towards understanding participatory processes: Framework, application and results, *J. Environ. Manage.*, 157, 84–95, 2015.
- Jahn, T., Bergmann, M., and Keil, F.: Transdisciplinarity between mainstreaming and marginalization, *Ecological Economics*, 79, 1–10, 2012.
- Jahn, T. and Keil, F.: An actor-specific guideline for quality assurance in transdisciplinary research, *Futures*, 65, 195–208, 2015.
- Kaufmann, A. and Kasztler, A.: Differences in publication and dissemination practices between disciplinary and transdisciplinary science and the consequences for research evaluation, *Science and Public Policy*, 36, 215–227, 2009.
- Klein, J. T.: Evaluation of interdisciplinary and transdisciplinary research, *Am. J. Prev. Med.*, 35, S116–123, 2008.
- Masse, L. C., Moser, R. P., Stokols, D., Taylor, B. K., Marcus, S. E., Morgan, G. D., Hall, K. L., Croyle, R. T., and Trochim, W. M.: Measuring collaboration and transdisciplinary integration in team science, *Am. J. Prev. Med.*, 35, S151–160, 2008.
- Mauser, W., Klepper, G., Rice, M., Schmalzbauer, B. S., Hackmann, H., Leemans, R., and Moore, H.: Transdisciplinary global change research: the co-creation of knowledge for sustainability, *Current Opinion in Environmental Sustainability*, 5, 420–431, 2013.
- OECD: Glossary of key terms in evaluation and results based management, available at: <http://www.oecd.org/development/peer-reviews/2754804.pdf> (last access: 14 July 2016), 2002.
- Pohl, C. and Hirsch Hadorn, G.: Principles for designing transdisciplinary research. A proposition by the Swiss Academies of Arts and Sciences, Oekom, München, Germany, 120 pp., 2007.

- Rafols, I., Laydesdorff, L., O'Hare, A., Nightingale, P., and Stirling, A.: How journal rankings can suppress interdisciplinary research, *Res. Policy*, 41, 1262–1282, <https://doi.org/10.1016/J.respol.2012.03.015>, 2012.
- Research Councils UK: Evaluation – Practical Guidelines, available at: <http://www.rcuk.ac.uk/documents/publications/evaluationguide-pdf/> (last access: 15 June 2017), 2011.
- Schuck-Zöllner, S., Bowyer, P., Jacob, D., and Brasseur, G.: Inter- und transdisziplinäres Arbeiten im Klimaschutz, in: *Anpassung an regionale Klimafolgen kommunizieren*, edited by: Beese, K., Fekkak, M., Katz, C., Körner, C., and Molitor, H., Oekom, 99–116, München, Germany, 2014.
- Schuck-Zöllner, S., Brinkmann, C., and Rödder, S.: Integrating Research and Practice – What climate services can learn from other fields, in: *Communicating climate change information for decision-making*, edited by: Serrao-Neumann, S., Coudrain, A., and Coulter, L., Springer, Heidelberg/New York, in press, 2018.
- Spaapen, J., Dijstelbloem, H., and Wamelink, F.: Evaluating research in context, consultative committee of sector councils for research and development (COS), The Hague, Netherlands, 156 pp., 2007.
- Stauffacher, M., Flüeler, T., Krütli, P., and Scholz, R. W.: Analytic and dynamic approach to collaboration: A transdisciplinary case study on sustainable landscape development in a Swiss prealpine region, *Syst. Pract. Action Res.*, 21, 409–422, <https://doi.org/10.1007/s11213-008-9107-7>, 2008.
- Vaughan, C. and Dessai, S.: Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework, *WIREs Clim. Change*, 5, 587–603, <https://doi.org/10.1002/wcc.290>, 2014.
- Wall, T. U., Meadow, A. M., and Horganic, A.: Developing evaluation indicators to improve the process of coproducing usable climate science, *Weather, Climate, and Society*, 9, 95–107, 2017.
- Walter, A., Helgenberger, S., Wiek, A., and Scholz, R.: Measuring societal effects of transdisciplinary research projects: Design and application of an evaluation method, *Eval. Program Plann.*, 30, 325–338, 2007.
- Wolf, B., Lindenthal, T., Szerencsis, M., Holbrook, J. B., and Heß, J.: Evaluating research beyond scientific impact, *GAIA*, 22/2, 104–114, 2013.