

Evaluating Research beyond Scientific Impact

How to Include Criteria for Productive Interactions and Impact on Practice and Society

Birge Wolf, Thomas Lindenthal, Manfred Szerencsits,
J. Britt Holbrook, Jürgen Heß

Concepts for the extension of research evaluation feature widely accorded criteria for activities and impacts of research. To enhance the use of existing criteria, reliable data must be compiled, with less effort for scientists and evaluators. Structuring existing reporting procedures for database input can increase usability.

Evaluating Research beyond Scientific Impact.

How to Include Criteria for Productive Interactions and Impact on Practice and Society

GAIA 22/2 (2013): 104 – 114

Abstract

Currently, established research evaluation focuses on scientific impact – that is, the impact of research on science itself. We discuss extending research evaluation to cover productive interactions and the impact of research on practice and society. The results are based on interviews with scientists from (organic) agriculture and a review of the literature on broader/social/societal impact assessment and the evaluation of interdisciplinary and transdisciplinary research. There is broad agreement about what activities and impacts of research are relevant for such an evaluation. However, the extension of research evaluation is hampered by a lack of easily usable data. To reduce the effort involved in data collection, the usability of existing documentation procedures (e. g., proposals and reports for research funding) needs to be increased.

We propose a structured database for the evaluation of scientists, projects, programmes and institutions, one that will require little additional effort beyond existing reporting requirements.

Keywords

data assessment, documentation, interdisciplinarity, organic agriculture, practice, productive interactions, research evaluation, social/societal impact, sustainability, transdisciplinarity

Problem Definition and Objectives: Extending Research Evaluation

Currently, established research evaluation procedures focus on scientific impact as measured by several proxies: publications, citation-based indicators, third-party funds, etc. This leads to disadvantages for interdisciplinary and transdisciplinary approaches, because evaluation influences reputation and research funding (Rafols et al. 2012). Furthermore, pressing problems such as global warming, food safety and the loss of ecosystem services (e. g., MA 2005) require that research activities – if not all, then at least those that are publicly funded – increase their contribution to solving these societally relevant problems and enhancing the applicability of findings, which often implies the involvement of non-academic target groups (e. g., Clark et al. 2005, VDW 2010, WBGU 2011 or Tåbara and Jaeger 2012). Hence, research evaluation is increasingly challenged to include criteria that go beyond mere scientific excellence (Spaapen and Van Drooge 2011). In the case of agricultural research evaluation there is a demand for criteria to cover interdisciplinarity and innovation transfer (DFG 2005) as well as social justice and ecological and social sustainability (see the *International Assessment of Agricultural Knowledge, Science and Technology for Development [IAASTD]* report by Albrecht 2009, p. 217). Thus, the main objective of our study *Development of a Manual for the Evaluation of Applied Research for Organic Agriculture (2008–2011)*¹ was to find procedures to extend research evaluation by adding criteria for productive interactions

Contact: Dipl.-Ing. Birge Wolf | Tel.: +49 5542 981536 |
E-Mail: birge.wolf@uni-kassel.de

Dipl.-Ing. Dr. Manfred Szerencsits | E-Mail:
mediator.szerencs@uni-kassel.de

Prof. Dr. Jürgen Heß | E-Mail: jh@uni-kassel.de

all: Department of Organic Farming and Cropping Systems |
University of Kassel | Nordbahnhofstr. 1a | 37213 Witzenhausen |
Germany

Dipl.-Ing. Dr. Thomas Lindenthal | University of Natural Resources and Life
Sciences, Vienna | Center for Global Change and Sustainability | Vienna |
Austria | E-Mail: thomas.lindenthal@boku.ac.at

Dr. J. Britt Holbrook | University of North Texas | Center for the Study of
Interdisciplinarity | Denton, Texas | USA | E-Mail: britt.holbrook@unt.edu

© 2013 B. Wolf et al.; licensee oekom verlag. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and the impact of research on practice and society,² with the focus on (organic) agricultural research. We also sought to identify barriers to such procedures. In the following, we present the main results of our study and propose a concept for improved data collection.

Material and Methods

Three core issues were investigated interdependently using three methodological approaches.

Core issues

1. *Review of the established evaluation system*, specifically from the perspective of applied, practice-oriented, interdisciplinary and transdisciplinary research (see Wolf et al. forthcoming).
2. *Reviews of existing concepts* for the evaluation of interdisciplinary and transdisciplinary research and social/societal/broader impact assessment.
3. *Suggestions* for the evaluation of practice-oriented organic agricultural research.

Methods

- *Literature analysis* of core issues 1 and 2.
- *Guideline-based qualitative interviews* with 22 agricultural researchers with regard to core issues 1 and 3: The scientists interviewed were selected to cover a broad range of subjects within (organic) agricultural sciences and neighbouring disciplines, to include expertise in interdisciplinary and transdisciplinary research, knowledge transfer, research management and research evaluation, and to involve a diversity of research institutions – universities, federal departments and

>

1 Funded by the German Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) in the *Federal Organic Farming Programme (BÖLN)*.

2 We refer to “productive interactions” because they comprise research practices that may lead to impact (see Spaapen and Van Drooge 2011) and the impact itself, with regard to both “practice” and “society”. We thus aim to cover the many terms used for evaluation beyond scientific impact, such as social, societal, broader, political and environmental impact, end-user relevance and the evaluation of inter- and transdisciplinary research. While “society” is the more inclusive definition, “practice” refers to the importance of the practical application of agricultural research. The term “practice orientation” includes all research activities providing results for use and benefit beyond science (practitioners and society) – regardless of the approach used.

Example for practice-oriented agricultural research: Below-root fertilisation with compost. The use of compost has a long history in sustainable food systems. Basic research discovered the biology of pathogen suppression by compost. Applied research gathered evidence of the effects in field trials. Together with an agricultural machinery maker, a machine for the line application of compost during potato planting was developed.



private or (semi-)public institutions. The interview guidelines included open questions relating to the core issues 1 and 3, and a semi-quantitative assessment of seven questions aimed at rating selected criteria and tools identified in the literature (on a scale from 1 “strongly agree”, to 6 “strongly disagree”). The interviews were transcribed, coded with *atlas.ti* and analysed using qualitative social research methods (Atteslander 2000, pp. 156–227, Flick 2002, p. 279).

- *Synthesis process*: workshop with interview partners, further discussions of results with agricultural scientists and representatives of research funding organisations; literature analysis relating to problems identified in documentation.

Evaluation beyond Scientific Impact in Different Fields

Concepts for an evaluation beyond scientific impact exist for different evaluation objects (project, programme, institution or scientist) and feature a high degree of conceptual diversity. We will focus mainly on ex post evaluation.

Many concepts are specifically tailored to certain approaches like interdisciplinary (e.g., Huutoniemi et al. 2010, Huutoniemi 2012) and transdisciplinary research (e.g., Pohl et al. 2011 for ex ante, Bergmann et al. 2005 for formative and ex post evaluation). Some are defined in combination with a certain discipline or research field and/or specification of the impact, for example, end-user relevance of applied agricultural research (Lyll et al. 2004), societal impact of translational research in medicine (Niederkröthaler et al. 2011), or impact of environmental research on policy (Shaw and Bell 2010).

Concepts that aim to evaluate the social/societal/broader impact of research in general are increasingly being developed, as evidenced by a special issue of *Research Evaluation* (Donovan 2011) and a recent literature review (Bornmann 2012). Key examples are the *Payback* model of the Health Economics Research Group (HERG), created for medicine and adapted for general use (Klautzer et al. 2011), and the *Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society (SIAMPI)*. The latter is based on the conception of *Evaluating Research in Context (ERiC)* (2010), part of the *Standard Evaluation Protocol* used by all Dutch universities and academic research organisations (Spaapen et al. 2011). Furthermore, the Research Councils UK (RCUK) insist on the inclusion of an “impact summary” and intended “pathways to impact” in proposals submitted (RCUK 2010), and on the ex post recording of outcomes and impacts for case studies (RCUK 2010, 2011a).

Another distinction can be drawn between pragmatic ad hoc concepts and theoretically grounded concepts. Examples of pragmatic concepts in agriculture are Formas (2007, 2009) and Pedersen et al. (2009), and some concepts of funding agencies that evaluate the impact of environmental research on policy (Shaw and Bell 2010, Bell et al. 2011). Theoretically grounded concepts are

based on certain implications as to how research should be conducted or how impact is generated. Evaluation of interdisciplinary and transdisciplinary research focuses on participative processes and knowledge integration (Nowotny et al. 2001, Pohl and Hirsch Hadorn 2006), which are important for change processes (Manring 2007) and the generation of target and transformation knowledge (Hennen et al. 2004, Pohl and Hirsch Hadorn 2006, p. 35). These qualitative concepts (table 1, p. 109) aim to contribute to learning processes in particular.

Some theoretical concepts use logic models³ divided into inputs, (processes), outputs, outcomes and impact. On the one hand, they are used in cost/benefit approaches, for example, in agri-environmental (Pearson et al. 2012) and in agricultural research in development cooperation, where impact is focused on improvements in the production of certain commodities (Davis et al. 2008). Pearson et al. (2012) as well as Davis et al. (2008) indicate shortcomings in the recording of societal and environmental impacts and their dependency on various assumptions. On the other hand, such logic models are used for qualitative evaluation like the HERG *Payback* model, where the authors point out that impact is a result of the whole system (Buxton 2011). This is also stated for agricultural innovation systems and applied in *Impact Pathway Evaluation* for research and innovation (Douthwaite et al. 2003) and in concepts for development cooperation (Reuber and Haas 2009).

Bridging processes and impacts without assessing the whole system is a task of the *SIAMPI* approach, another theoretical grounded model, which focuses on productive interactions as a prerequisite and proxy for impact. Productive interactions are defined here as “exchanges between researchers and stakeholders in which knowledge is produced and valued that is both scientifically robust and socially relevant” (Spaapen and Van Drooge 2011, p. 212). They are divided into direct, indirect (by publications, exhibitions, etc.) and financial interactions (Spaapen and Van Drooge 2011). The *SIAMPI* concept – in combination with *Impact Pathway Evaluation* – was also used by the French National Institute for Agricultural Research (INRA) to evaluate cases where sets of actions related to certain impacts (ASIRPA 2012).

Results: Criteria, Evaluation Tools and Challenges for the Extension of Research Evaluation

The results of interviews, synthesis workshop and literature analysis presented in this paper concern the following aspects:

- criteria for the evaluation of practice-oriented (organic) agricultural research;
- evaluation tools;
- challenges confronting the extension of research evaluation.

³ Logic models are in general ideal-typical distinctions. As Spaapen et al. (2011) have stated, in reality the categories often overlap.

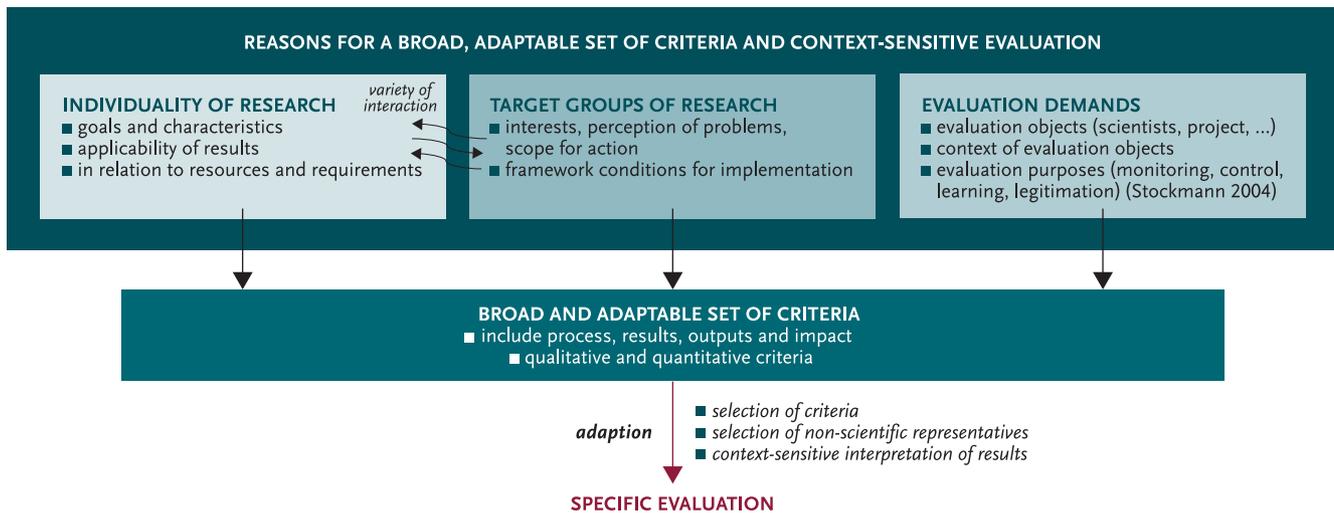


FIGURE 1: Interview results concerning the evaluation of practice-oriented research in (organic) agriculture: interviewees and workshop participants favour specific evaluation on the basis of a broad and adaptable set of criteria.

Criteria for the Evaluation of Practice-oriented (Organic) Agricultural Research

The main results of the interviews are a clear statement indicating the necessity for a broad and adaptable set of criteria, the identification and rating of core issues for evaluation criteria, and the suggestion of criteria for productive interactions and the impact of research on practice and society.

The interview results are followed by the number of persons who offered suggestions in relation to the total number of interviewees⁴ (e.g., 6/22) or participants of the synthesis workshop (e.g., 4/10 WS). The outcome of the semi-quantitative assessment is marked with the initial (Q).

Necessity for a Broad, Adaptable Set of Criteria

Figure 1 shows the interviewees' opinion that evaluation of practice-oriented research in (organic) agriculture should be based on a broad set of criteria (11/22 + 7/10 WS). Broad means 1. research practices that may lead to impact (productive interactions) as well as the subsequent impact itself (11/22) and 2. both quantitative and qualitative criteria (7/22). Interviewees argue in favour of a broad, adaptable set of criteria so as to respect the individuality of research and differences in target groups, leading to unique impacts and specific benefits. According to the goal of our study – i.e., extending research evaluation in general – the criteria need to fit and be adaptable to specific evaluation objects, contexts, purposes and times. The process of adaption should include and be open to the selection of criteria (4/22 + 7/10 WS), the selection and integration of representatives from practice and/or society (8/21) as well as the consideration of specific conditions for scientists' work, for example, in relation to resources and

requirements for research (2/10 WS) or the context for implementation in practice (8/22).

The preference observed here for a broad set of criteria corresponds to recommendations in the evaluation literature (Holbrook and Frodeman 2011, Frodeman and Holbrook 2011). Quality is seen as a relative concept “driven by the variability of goals and criteria” (Klein 2006, p.75), and the combination of narratives with relevant qualitative and quantitative indicators is regarded as state-of-the-art (Donovan 2011). Criteria and interpretation of evaluation results can be adapted according to the characteristics and contexts of projects (Daschkeit and Loibl 2007, ERiC 2010) or the goals of research programmes (Braun et al. 2009), or developed with the participation of evaluatees (Stokols et al. 2003, Bergmann et al. 2005, Blackstock et al. 2007) or evaluatees and practitioners (Pedersen et al. 2009).

Core Issues for Evaluation from Scientists' Viewpoints

Following the qualitative analysis (figure 2, p. 108), interviewees consider practice orientation and interdisciplinarity to be of particular importance. They also discussed intensively social relevance, as it was found to be important but difficult to assess. Diversity of topics seemed to be an issue primarily at research funding level. Other issues were often seen “in conjunction with” or “in relation to” practice impact and interdisciplinarity. The majority of the interviewed scientists pointed out that target groups and other persons and groups affected by research need to be identified and taken into account (17/21), especially for the evaluation of social/societal relevance and sustainability.

Criteria for the Evaluation of Impact on Practice and Society

In accordance with the results from the interviews, we concentrate here on criteria relating to productive interactions and their subsequent impact on practice and society. The criteria (table 2, p. 110) have been drawn from the literature review and the results

⁴ The total number varies, as some questions remained unanswered by a few interviewees due to lack of time.

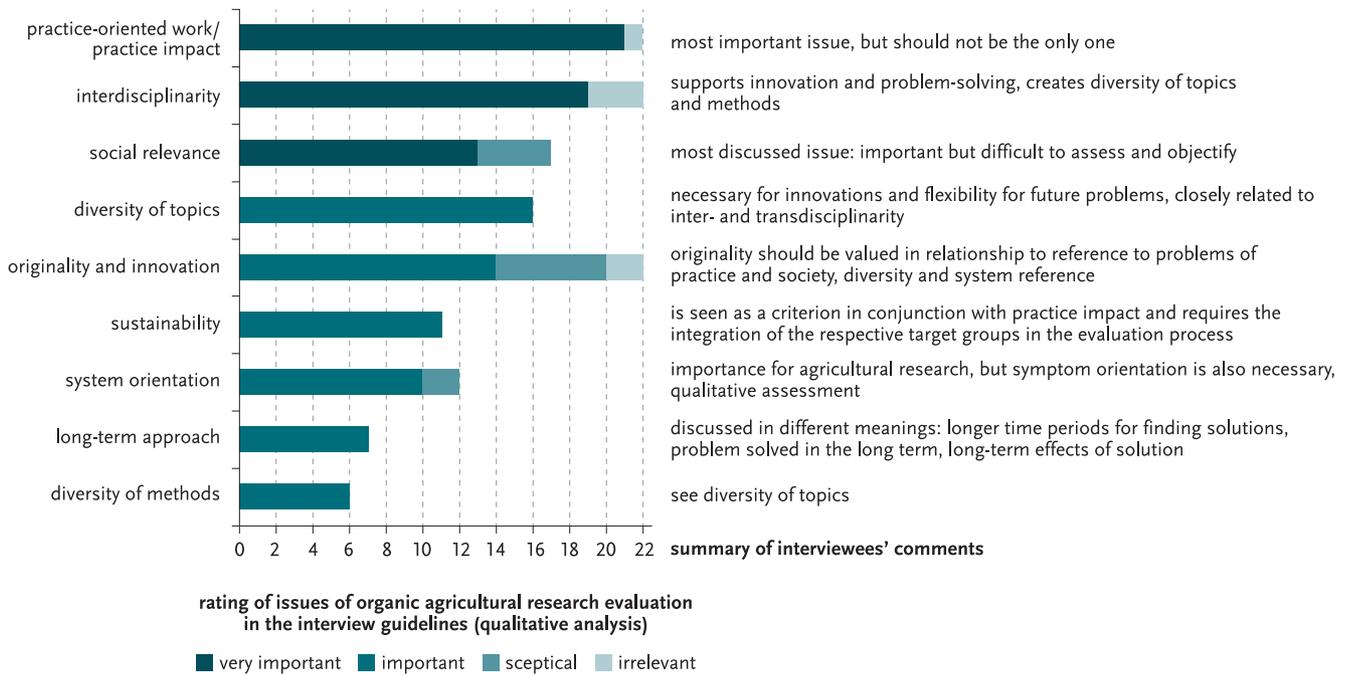


FIGURE 2: Core issues of organic agricultural research evaluation, which were asked for in the interview guidelines. Ratings derived from qualitative analysis of the interviews.

of the interviews and synthesis workshop. On the basis of our literature review, we selected concepts for the detailed analysis of criteria and evaluation tools, with the focus on formative and ex post evaluation and a broad range of evaluation objects (table 1). In table 2, we cite the concepts with the numbers used in table 1 if they refer to similar criteria or involve criteria that seemed suitable for supplementation of the interview results.

The criteria listed in table 2 can be used for the interdependent evaluation of scientists, projects, institutions and research programmes. They are characterised on a three-level scale referring to their rough estimated applicability.

The concepts differ in extent and detail. Having said that, table 2 shows a high level of agreement within the literature and interview results regarding the criteria proposed. In general, processes and outputs receive more emphasis in the literature than impact. Interviewees mention content, contacts, publications and products as preconditions for impact and provide detailed criteria, most of which seem to be applicable with moderate effort. Suggestions for impact assessment – seen as outstandingly important for agricultural research – are underlined with less detailed criteria, but can be divided into three main stages: response, application and impact of application. There are also fewer concepts with criteria for impact because of the specific nature of the challenges involved in impact assessment.

Evaluation Tools

Evaluation tools and methods were reviewed in the literature (below cited with the numbers from table 1) and explored in the interviews. We focus on the usability of tools for the extension of

research evaluation to include productive interactions and the impact of research on practice and society.

There are various well defined tools for external and internal evaluation, which are usually applied using multi-method approaches and triangulation (11, 14, 16, 20, see also Boaz et al. 2009, Donovan 2011). Mostly they include peer review and/or panels of scientists (10, 11, 12, 17, 18). In some cases practitioners or societal stakeholders contribute to reviews (1) and panels (1, 2, 5, 10, 11), as was also suggested in the interviews (Q 15/16). Discursive evaluation⁵ approaches are advocated in literature (11, 12, 13, 15, 20) and interviews (Q 15/16), and some of the concepts use monitoring or process supervision as evaluation tools during research (10, 14, 19).

Most evaluations are carried out as “stand-alone procedures”, which means that data are assessed solely for one specific evaluation. Thus existing sources like project reports (1, 3, 8, 10, 15, 16, 18) or data/documentation of research programmes (8, 10, 18) are used. Usually, additional information is collected via interviews with the evaluatees (2, 3, 8, 10, 15, 16, 18) and/or scientific experts (2, 8, 10, 11), and in some cases via workshops/focus groups or mentoring (10, 13, 18). Some evaluations also include data collection within practice and society with interviews (2, 3, 5, 8, 15) or workshops (3, 10). With the focus on assessing the impact on policy, Boaz et al. (2009) also show the predominant use of interviews, case studies and documentary analysis. Data collection thus re-

⁵ Discursive evaluation is a process in which evaluatees and evaluators reflect on evaluation results in a constructive discourse (cf. Bergmann et al. 2005).

TABLE 1: For the detailed analysis of evaluation criteria and tools, 18 concepts were selected with the focus on formative and ex post evaluation in transdisciplinary and agricultural research and societal impact in general.

selected concepts	object of research evaluation
agricultural	
1 Formas (2009)	project
2 Pedersen et al. (2009)	project
3 Lyall et al. (2004)	institute/programme
4 Davis et al. (2008)	project
general	
5 ERiC (2010)	institution/research group
6 SIAMPI: Spaapen and Van Drooge (2011)	projects/institutions
7 RCUK (2010, 2011 a, 2011 b)	projects/programme and service for institutions
medical and general	
8 Payback: Klautzer et al. (2011)	project
9 Niederkrotenthaler et al. (2011)	addition for scientific publications
inter- and transdisciplinary	
10 Begusch-Pfefferkorn (2006)	programme
11 Defila et al. (2008, pp. 72–73)	research group
12 Stoll-Kleemann (2007)	project
13 Schiller et al. (2005)	project/research group/department
14 Bergmann et al. (2005)	project
15 Blackstock et al. (2007)	project
16 Defila and Di Giulio (1999)	project
17 Schübel (2007)	project
18 Walter et al. (2007)	project, programme

quires great effort. One example of generalised continuous data assessment can be taken from the RCUK (7). The *Outcome Collection System* uses a database to record outcomes and impact for three years after project completion (RCUK 2010, 2011 a). The system also provides data for, and accommodates reporting requirements of the *Research Excellence Framework* and the Higher Education Statistics Agency (RCUK 2010, 2011 b).

Interviewees mentioned the same possible data sources for “stand-alone procedures”, but put more emphasis on “self-documentation” of scientists’ productive interactions and impacts as a basis for the general extension of research evaluation (9/10 WS). Core issues of database-assisted self-documentation were that it needs to be complemented by feedback from practice and society and that additional effort relating to documentation by scientists needs to be avoided.

Challenges to Extending Research Evaluation

The direct evaluation of impact is most challenging. It raises the question of the attribution of impact to research, which in “systems of innovation” is regarded more in terms of “contribution”. Additionally, it is difficult to deal with the time gap between research and impact, which includes unexpected and unintended impacts (Buxton 2011, Spaapen and Van Drooge 2011). Therefore,

the evaluation of preconditions for impact, such as processes and outputs, is the predominant focus in evaluation concepts.

However, evaluation beyond scientific impact is generally confronted with problems regarding the way scientific and non-scientific reviewers are selected and whether or not they are sufficiently competent (e.g., Daschkeit 2007, Skolits et al. 2009). Furthermore, shortcomings in the extent, availability and quality of data (e.g., Uriarte et al. 2007, Spaapen and Van Drooge 2011, Bell et al. 2011, Holbrook and Frodeman 2011) were extensively discussed in the literature and interviews, thus becoming a point of focus in our further investigation.

Criteria beyond scientific impact are already used for ex ante assessment, for example, in specific calls of the *Seventh Framework Programme* (Regulation [EC] 1906/2006, Holbrook and Frodeman 2011) and grant proposals for the US National Science Foundation (NSF) (Frodeman and Holbrook 2011, Holbrook 2012). Braun et al. (2009) remark that data available for the *Framework Programmes* are more suitable for programme administration than for ex post evaluation. They recommend simplified databases for the creation of indicators. In the *Seventh Framework Programme*, for example, contributions to sustainable devel-

opment based on expected impacts were monitored via a qualitative text analysis and interviews (Martinuzzi and Hametner 2012).

In addition, work is being done on the standardisation of data collection at institutional level (e.g., EC 2010, Van Vught and Ziegele 2011, *Science and Technology for America’s Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science [STAR METRICS]*⁶, WR 2013). However, attempts to assess the transfer of knowledge, social impact or regional engagement are hampered by a lack of reliable data (EC 2010, Van Vught and Ziegele 2011).

A Concept for Improved Data Collection

The lack of easily usable data for evaluation beyond scientific impact is a problem affecting all evaluation objects from project to programme, from scientists to institutions. Accordingly, we developed an initial concept for continuous data collection that is independent of individual evaluations but provides a basic serv-

>

6 For more information see STAR METRICS project homepage [www.starmetrics.nih.gov/Star/Participate#about\[0\]](http://www.starmetrics.nih.gov/Star/Participate#about[0]).

TABLE 2: Criteria for evaluating practice-oriented achievements of scientists, projects, institutions and programmes.

category (number of mentions)	important sub-criteria (number of mentions)	applicability ^a	selected concepts (table 1) ^b mentioning similar criteria and criteria to be supplemented from the literature
content: potential relevance for practice and society 13/22 ^c +10/10 WS ^d	<ul style="list-style-type: none"> ■ scientific questions consider problems of practitioners or society (6) ■ knowledge of stakeholders included in research process (1+10 WS) 	B, C	3, 12, 14, 16, 18: general presumption for transdisciplinarity; adequately broad know-how for problem-solving, in particular relating to causes of problems, natural/economic/judicial connections, feasibility 13
	<ul style="list-style-type: none"> ■ contribution to problem-solving (8) ■ feedback from practice/society (12), e.g., applicability (12), potential benefit (8), and risks (3) 	B, C	1, 2, 12, 14, 15, 16, 18: interest of stakeholders 5; strategy for commitment/support of application 14, 16
contacts with actors in practice and society/practice-oriented research process 14/22+5/10 WS ^d	<ul style="list-style-type: none"> ■ contacts in general (14 + 5WS) ■ dissemination/presentations (9) ■ communication/contact/collaboration (incl. workshops, field days)(8) (Q 16/16^e) ■ networking (Q 14/16), capacity building (3) ■ contact/collaboration supported with structures and methods (9) ■ quality of contact/collaboration (8) ■ learning process/synthesis between scientific and practical knowledge (5) ■ co-financing by practice (2 advocating, 1 refusing) 	A, B A A, B A, B B, C B, C A, B	direct productive interactions 6; translation into practice undertaken 9 5, 6, 7, 8, 10, 12, 13, 14, 16 3, 5, 6, 7, 9, 10, 12, number of non-scientific experts in projects 5, 10 4, 5, 6, 7, 9, 10, 14, 15, 18 1, 2, 6, 14, 16: representative integration of different views 15 3, 6, 12, 14, 16, 18: internal/external transparency of project 15 5, 6, 7, 10, 13: financial productive interactions 6
publications and products for practice (and society) 13/22 + 5/10 WS	<ul style="list-style-type: none"> ■ articles in newspapers, periodicals, trade journals, magazines (9), specialist books (2), electronic media (2), recommendations for advisory service (Q 14/15) ■ products (1) ■ preparation of results and media appropriate for target groups (8+2 WS) 	A A, B C	publications 3, 5, 6, 7, 10, 12, 13, 14, 16: indirect productive interactions 6; exhibition/performance 7; products 3, 6, 8, 14, 16; patents; trade mark 7; outputs/innovation: technical, political, capacity-building 4; many different outputs 7 1, 2, 3, 6, 14, 16
impact of research facts about response, application, and problem-solving (follow-up assessment) 21/22 + 2/10 WS	<ul style="list-style-type: none"> ■ <i>response in practice and society</i> ■ reaching relevant target groups/multipliers (4+2 WS) ■ size/ characteristics of target groups (5+2 WS) ■ follow-up projects with practice (1) <p>(adapted) application of research results, concepts, products (20)</p> <ul style="list-style-type: none"> ■ extent of use: user groups (5+2 WS), area (1), sales volume, etc. (3) ■ impact on politics (Q 12/16) 	B, C B, C B, C	reaching relevant target groups, 5, 14: national/international awards from practice, education or politics 7, 10, 13; occupational change to practice 13 3, 5, 8, 12, 14, 16, 18: impact on politics 7, 8, 10: ratio: de facto/potential user group 4; translation into practice accomplished: level (regional, national, international), status (preliminary, permanent), target group (individuals, subpopulations, public) 9; spinoff company, exploitation 7 5, 8, 14, 15: extent of what is the problem solved 14, 16; cost/benefit ratio of impact 4; sustainability categories 1, 4; detailed impact summary, various impact types 7
	<ul style="list-style-type: none"> ■ <i>impact of application/contribution to problem solving</i> (7), e.g., (extent of) benefit (11) for target group, environment and society resp. in sustainability categories 	B, C	

a A: quantitative, easily applicable; B: quantitative and qualitative assessment combined, more effort involved; C: involvement of non-scientific persons required, with the amount of effort involved depending on the method used | b numbers refer to selected concepts in table 1 | c number of suggestions in the interviews/total number of interviewees | d number of suggestions in the workshop/total number of workshop participants (WS) | e outcome of the semi-quantitative assessment (Q)

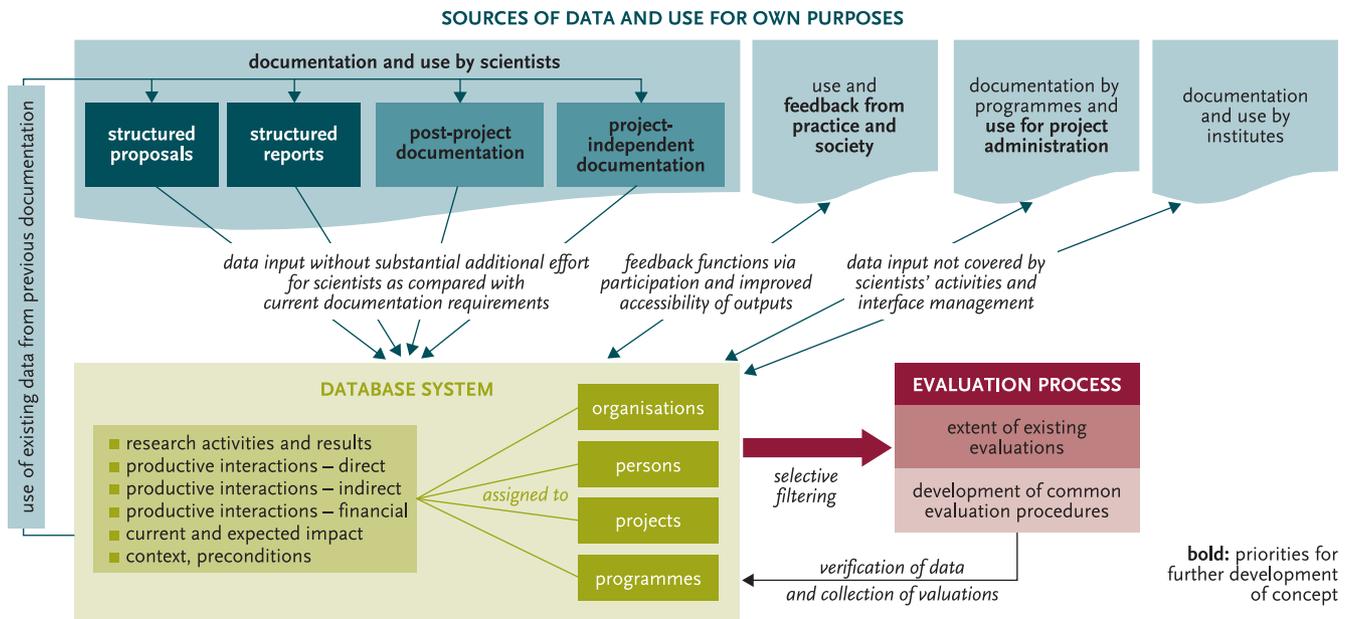


FIGURE 3: Improved data collection: The use of structured proposals and reports (and other existing data sources) for the database system reduces the amount of effort required. The database allows data to be assigned to different evaluation objects and used for different evaluation processes.

ice for them. It aims to reduce the time and expense involved, and enables existing evaluation processes (e.g., in institutes) to be extended and common evaluation procedures for practical and societal impact to be developed (e.g., in conjunction with scientific publications like suggested by Niederkrotenthaler et al. 2011 or an ex post peer review of projects).

The concept outlined in figure 3 shows that the main data sources are structured proposals and reports. They include information about research processes, results, productive interactions, impact and context. Via the database system they can be assigned to and filtered for different evaluation objects and serve evaluation processes.

In the following, we describe the concept in terms of its required characteristics: Documentation should

- **be broadly applicable** because it 1. transcends the borders of individual evaluation objects, 2. includes different types of research, for example, different approaches, topics and disciplines, and 3. accommodates multiple evaluation purposes and a complete range of productive interactions and impacts. A broadly applicable documentation enables connections to be made between the establishment of an extended evaluation and the high variety of objectives of evaluation.
- **use application and reporting for projects⁷ in a structured form (example in figure 4, p. 112) as the main (but not the only!) source of data.** Most research is conducted in the context of projects, which have to be documented anyway. Therefore, the type of documentation may change, but the amount of effort involved in it does not increase. Using structured applications and reports also provides benefits for research funding by facilitating project administration and further development. Furthermore, use by research funding will assure the participation of

a critical mass of scientists for a profound and regularly updated dataset (see Holbrook 2012).

- **enable data input whenever results are generated or impacts observed** – independent of any specific evaluation. This will alleviate the time gap problem.
- **support the assignment of productive interactions and impacts to different evaluation objects** (project, programme, institution or scientist) via a database. The combination of categorised and free text information ensures 1. high usability for different evaluation objects, purposes and research contexts via individual filtering and detailed or aggregated use of data, 2. transparency and reliability of data, and 3. consideration of the individuality of research.
- **be grounded in the “reality” of practice-oriented research** to ensure robust, reliable data. Thus, documentation needs to be developed with and subsequently driven by researchers, and should be complemented by information from practice and society, in order to record the real productive interactions and impacts and admit of more viewpoints in evaluation than the purely scientific ones.

For future developments the benefit can be broadened or linked with current trends in science, like open access and potentialities of new media, and possibilities for stakeholder communication. >

7 A rough analysis of requirements for research project applications and reports used by the German government departments, together with consideration of the objectives, documented for the BMELV research in general and its programmes BÖLN and Innovationsförderung, indicates that they go a long way toward covering the information needs of evaluation but – as already mentioned – not as easily usable data (see Wolf et al. forthcoming).

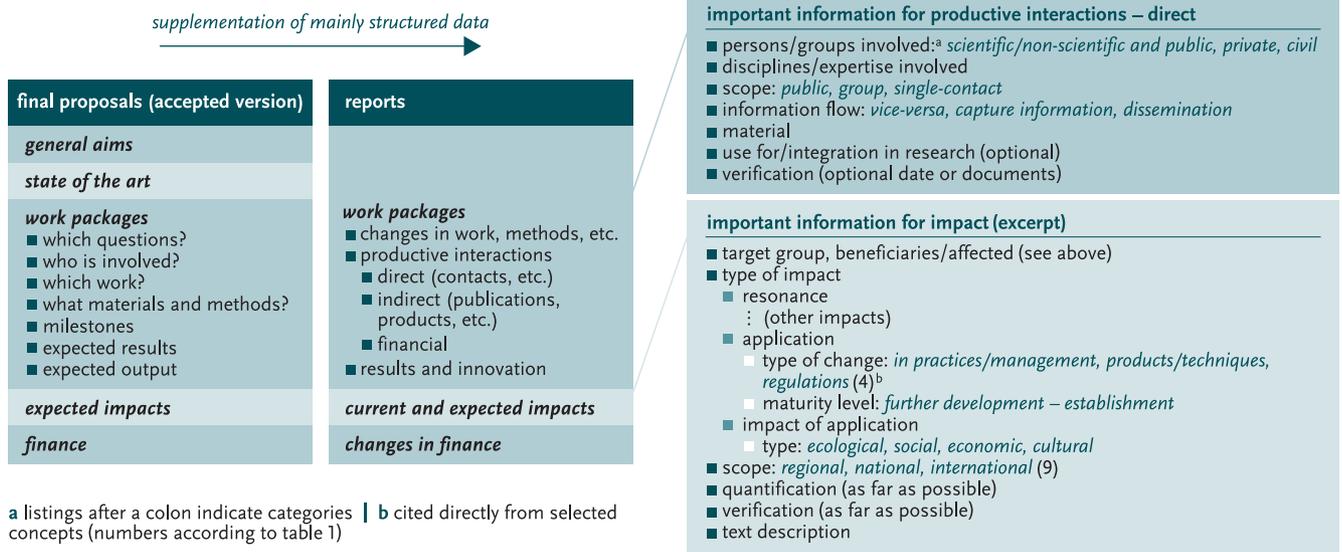


FIGURE 4: First draft of the structure of proposals and reports (left) as an input for the database system shown in figure 3. Exemplified details of information for productive interactions and impact, accessed in the structured reports via text and categories (right).

The overview of documentation in figure 3 is detailed in figure 4. It illustrates the possible structure of proposals and reports and gives an example of the information that should be accessed for direct productive interactions and impact.

Discussion and Conclusions: Improved Documentation and Commitment in the Science-Society System Are Needed

The proposed concept for improved data collection is based on the variety of reviewed evaluation concepts and their substantial overlap in terms of criteria. Accordingly, even evaluations with different foci may be based on the same data – and all are currently hampered by the lack of reliable, easily usable data.

Our study started with the focus on agricultural research, but the literature analysis shows comparable challenges in different research fields. Therefore, our findings seem to be applicable in other disciplines too. A categorisation based on the properties instead of the “names” of productive interactions and impacts (e. g., vice-versa communication, practical demonstration and target group instead of “field day”) (figure 4) will serve this generalisation.

The extension of evaluation beyond scientific impact requires discussion by a wide range of actors in the science-society system, which is already in progress. Indeed, in our study the interviews and literature review of criteria were carried out independently. For that reason, the interviews deliver important background information for developing criteria, tools and data assessment, but lack a discussion of existing criteria and tools. This discussion will be conducted in the follow-up project with regard to the data available.

The documentation procedure described above will be further developed in a test phase, integrating the needs of scientists, evaluators and project funding organisations. We will also explore the inclusion of non-scientific persons in the evaluation process. Our further research will aim at a high degree of procedure usability and verify whether or not the expected benefits can be achieved.

Using the data for different purposes, for example, funding decisions and learning processes (e. g., Rogers and Jordan 2010, Manring 2007) would indeed raise questions of trust and users’ conflicting interests. Thus, verification, careful use of data and evaluation consequences, and/or an accepted institution to run the database would be required.

Some might see extended evaluation as a threat to the “freedom of research”, but it might equally well be asked whether the established hierarchical and discipline-based evaluation system really serves that freedom. From our point of view, evaluation with regard to scientific excellence and societal benefits opens up freedom in terms of the plurality of research (cf. Frodeman et al. 2012), and strengthens the democratisation of public research funding as called for by the Federation of German Scientists (VDW 2010).

We would like to thank the *Federal Organic Farming Programme (BÖLN)* of the German Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) for its financial support – and all those actively involved in the study as well as the audiences for their constructive discussion and advice.

References

- Albrecht, S. 2009. *Weltagrabericht. Synthesebericht*. Hamburg: Hamburg University Press.
- ASIRPA (Socio-economic Analysis of Impacts of Public Agronomic Research). 2012. *ASIRPA project in 4 pages*. Paris: French National Institute for Agricultural Research (INRA). www6.inra.fr/asirpa_eng/ASIRPA-project/International-ASIRPA-workshop-Nov-2012/General-material (accessed February 11, 2013).

- Atteslander, P. 2000. *Methoden der empirischen Sozialforschung*. 9th edition. Berlin: Walter de Gruyter.
- Begusch-Pfefferkorn, K. 2006. KLF-Wissensbilanz. Leistungsdarstellung des inter- und transdisziplinären Programms Kulturlandschaftsforschung. In: *Fachübergreifende Nachhaltigkeitsforschung. Stand und Visionen am Beispiel nationaler und internationaler Forscherverbände*. Edited by B. Glaeser. Munich: oekom. 119–139.
- Bell, S., B. Shaw, A. Boaz. 2011. Real-world approaches to assessing the impact of environmental research on policy. *Research Evaluation* 20/3: 227–237.
- Bergmann, M. et al. 2005. *Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten*. Frankfurt am Main: Institut für sozial-ökologische Forschung (ISOE).
- Blackstock, K. L., G. J. Kelly, B. L. Horsey. 2007. Developing and applying a framework to evaluate participatory research for sustainability. *Ecological Economics* 60/4: 726–742.
- Boaz, A., S. Fitzpatrick, B. Shaw. 2009. Assessing the impact of research on policy: A literature review. *Science and Public Policy* 36/4: 255–270.
- Bornmann, L. 2012. What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology* 64/2: 217–233.
- Braun, M. et al. 2009. *Tools and indicators for community research evaluation and monitoring*. Bad Camberg: Proneos. http://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/evaluation_studies_and_reports/evaluation_studies_and_reports_2009/tools_and_indicators_for_community_research_evaluation_and_monitoring.pdf (accessed October 24, 2012).
- Buxton, M. 2011. The payback of “payback”: Challenges in assessing research impact. *Research Evaluation* 20/3: 259–260.
- Clark, W. C., P. J. Crutzen, H. J. Schellnhuber. 2005. *Science for global sustainability: Toward a new paradigm*. CID Working Paper 120. Cambridge, MA: Science, Environment and Development Group, Center for International Development, Harvard University.
- Daschkeit, A. 2007. Zur Beurteilung transdisziplinärer Forschung. Hinweise auf Bücher zur Transdisziplinarität. *GAIA* 16/1: 58–65.
- Daschkeit, A., M. C. Loibl. 2007. Projektspezifische Evaluation human-ökologischer Nachhaltigkeitsforschung. Entwurf eines Bewertungsschemas. In: *Evaluation inter- und transdisziplinärer Forschung. Humanökologie und Nachhaltigkeitsforschung auf dem Prüfstand*. Edited by S. Stoll-Kleemann, C. Pohl. Munich: oekom. 157–171.
- Davis, J., J. P. D. Gordon, D. Templeton. 2008. *Guidelines for assessing the impacts of ACIAR's research activities*. ACIAR Impact Assessment Series Report 58. Canberra: Australian Centre for International Agricultural Research (ACIAR).
- Defila, R., A. Di Giulio. 1999. Evaluationskriterien für inter- und transdisziplinäre Forschung. *Panorama* 99: 5–11. www.ikaoe.unibe.ch/forschung/ip/Sondernummer.Pano.1.99.pdf (accessed January 26, 2010).
- Defila, R., A. Di Giulio, M. Scheuermann. 2008. *Management von Forschungsverbänden. Möglichkeiten der Professionalisierung und Unterstützung. Standpunkte*. Weinheim: Wiley-VCH.
- DFG (German Research Foundation). 2005. *Perspektiven der agrarwissenschaftlichen Forschung/Future Perspectives of Agricultural Science and Research. Denkschrift/Memorandum*. Weinheim: Wiley-VCH.
- Donovan, C. 2011. State of the art in assessing research impact: Introduction to a special issue. *Research Evaluation* 20/3: 175–179.
- Douthwaite, B., T. Kuby, E. van de Fliert, S. Schulz. 2003. Impact pathway evaluation: An approach for achieving and attributing impact in complex systems. *Agricultural Systems* 78: 243–265.
- EC (European Commission). 2010. *Assessing Europe's university-based research: Expert group of assessment of university-based research*. EUR 24187. Luxembourg: Publications Office of the European Union. http://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf (accessed July 4, 2011).
- ERIC (Evaluating Research in Context). 2010. *Evaluating the societal relevance of academic research: A guide*. ERIC publication 1001 EN. www.siampi.eu/Content/SIAMP/ERIC%20Guide%202010.pdf (accessed January 26, 2010).
- Flick, U. 2002. *Qualitative Forschung. Ein Handbuch*. 2nd edition. Reinbek: Rowohlt.
- Formas (Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning). 2007. *Evaluation report 2006: Evaluation of research on organic production in Sweden*. Stockholm: EDITA Ljunglöfs.
- Formas. 2009. *Evaluation report 2009: Evaluation of SLF-funded research 2002–2007*. Stockholm: Forskningsrådet Formas. www.formas.se/PageFiles/5324/Formas_SLF.pdf (accessed May 7, 2013).
- Frodeman, R., J. B. Holbrook. 2011. NSF's struggle to articulate relevance. *Science* 333/6039: 157–158.
- Frodeman, R., J. B. Holbrook, C. Mitcham, H. Xiaonan. 2012. *Peer review, research integrity, and the governance of science: Practice, theory, and current discussions*. Beijing: People's Publishing House.
- Hennen, L., T. Petermann, C. Scherz. 2004. *Partizipative Verfahren der Technikfolgen-Abschätzung und parlamentarische Politikberatung. Neue Formen der Kommunikation zwischen Wissenschaft, Politik und Öffentlichkeit*. TAB-Arbeitsbericht 96. Berlin: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB).
- Holbrook, J. B. 2012. Re-assessing the science-society relation: The case of the US National Science Foundation's broader impacts merit review criterion (1997–2011). In: *Peer review, research integrity, and the governance of science: Practice, theory, and current discussions*. Edited by R. Frodeman, J. B. Holbrook, C. Mitcham, H. Xiaonan. Beijing: People's Publishing House. 328–362.
- Holbrook, J. B., R. Frodeman. 2011. Peer review and the ex ante assessment of societal impacts. *Research Evaluation* 20/3: 239–246.
- Huutoniemi, K. 2012. *Interdisciplinary accountability in the evaluation of research proposals: Prospects for academic quality control across disciplinary boundaries*. Dissertation, University of Helsinki, Finland.
- Huutoniemi, K., J. T. Klein, H. Bruun, J. Hukkinen. 2010. Analyzing interdisciplinarity: Typology and indicators. *Research Policy* 39/1: 79–88.
- Klautzer, L., S. Hanney, E. Nason, J. Rubin, J. Grant, S. Wooding. 2011. Assessing policy and practice impacts of social science research: The application of the payback framework to assess the *Future of Work* programme. *Research Evaluation* 20/3: 201–209.
- Klein, J. T. 2006. Afterword: The emergent literature on interdisciplinary and transdisciplinary research evaluation. *Research Evaluation* 15/1: 75–80.
- Lyall, C., A. Bruce, J. Firn, M. Firn, J. Tait. 2004. Assessing end-use relevance of public sector research organisations. *Research Policy* 33/1: 73–87.
- MA (Millennium Ecosystem Assessment). 2005. *Ecosystems and human well-being: Synthesis*. Washington, D. C.: Island.
- Manring, S. L. 2007. Creating and managing interorganizational learning networks to achieve sustainable ecosystem management. *Organization and Environment* 20/3: 325–346.
- Martinuzzi, A., M. Hametner. 2012. Monitoring expected impacts of the *EU Framework Programme* on sustainable development. *10th Biennial Conference of the European Evaluation Society*, October 1–5, Helsinki, Finland.
- Niederkröthaler, T., T. E. Dorner, M. Maier. 2011. Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists. *BMC Public Health* 11/1: 588–596.
- Nowotny, H., P. Scott, M. Gibbons. 2001. *Re-thinking science: Knowledge and the public in an age of uncertainty*. Cambridge, UK: Polity.
- Pearson, S., P. Chudleigh, S. Simpson, N. Schofield. 2012. Learning to invest better: Using ex post investment analysis on agri-environmental research and development. *Research Evaluation* 21/2: 136–151.
- Pedersen, S. M., M. V. Boesen, D. L. A. Baker. 2009. A model for evaluating food and agricultural research projects: A Danish application. In: *Proceedings of the 17th International Farm Management Congress: Agriculture: Food, fiber and energy for the future*. Edited by H. H. Guither, J. L. Merry, C. E. Merry. Bloomington, IL: International Farm Management Association (IFMA). 99–114.
- Pohl, C. et al. 2011. *Questions to evaluate inter- and transdisciplinary research proposals*. Working paper. Bern: td-net for Transdisciplinary Research. www.transdisciplinarity.ch/documents/td-Evaluation2011_workingpaper.pdf (accessed May 6, 2013).
- Pohl, C., G. Hirsch Hadorn. 2006. *Gestaltungsprinzipien für die transdisziplinäre Forschung*. Munich: oekom.
- Rafols, I., L. Leydesdorff, A. O'Hare, P. Nightingale, A. Stirling. 2012. How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy* 41/7: 1262–1282.

>

- RCUK (Research Councils UK). 2010. *Research outcomes project invitation to tender research outcome types*. Swindon, UK: RCUK. www.rcuk.ac.uk/documents/oocp/OutputTypes.pdf (accessed September 21, 2011).
- RCUK. 2011a. *RCUK impact requirements: Frequently asked questions*. Swindon, UK: RCUK. www.rcuk.ac.uk/documents/impacts/RCUKImpactFAQ.pdf (accessed September 21, 2011).
- RCUK. 2011b. *RCUK research outcomes system now available*. Swindon, UK: RCUK. www.rcuk.ac.uk/media/news/2011news/Pages/111124.aspx (accessed October 24, 2012).
- Regulation (EC) No 1906/2006. 2006. *Regulation (EC) No 1906/2006 of the European Parliament and of the Council of 18 December 2006 laying down the rules for the participation of undertakings, research centres and universities in actions under the Seventh Framework Programme and for the dissemination of research results (2007–2013)*. Official Journal of the European Union L 391: 1–18.
- Reuber, M., O. Haas. 2009. Evaluierungen in der Deutschen Gesellschaft für Technische Zusammenarbeit (GTZ). *Zeitschrift für Evaluation* 6/1: 137–150.
- Rogers, J. D., G. Jordan. 2010. New research evaluation frameworks and methods for systems level learning: Introduction to a special section. *Research Evaluation* 19/4: 235–237.
- Schiller, J. et al. 2005. *Evaluation interdisziplinärer Programm-Forschung. Ziele, Anlässe und Prozesse*. UFZ-Diskussionspapiere 17/2005. Leipzig: Helmholtz-Zentrum für Umweltforschung – UFZ.
- Schübel, H. R. 2007. Interdisziplinäre Kooperation in transdisziplinärer Forschung besser machen? Prozessbegleitung als formative Evaluation. In: *Evaluation inter- und transdisziplinärer Forschung. Humanökologie und Nachhaltigkeitsforschung auf dem Prüfstand*. Edited by S. Stoll-Kleemann, C. Pohl. Munich: oekom. 195–210.
- Shaw, B., S. Bell. 2010. *How to evaluate the impact of environmental research on policy: Guidelines and supporting information*. London: Policy Studies Institute. www.psi.org.uk/publications/SKEP/Guidelines%20Final%20v1.2.pdf (accessed August 22, 2012).
- Skolits, G. J., J. A. Morrow, E. M. Burr. 2009. Reconceptualizing evaluator roles. *American Journal of Evaluation* 30/3: 275–295.
- Spaapen, J. et al. 2011. *SIAMPI final report*. www.siampi.eu/Content/SIAMPI/SIAMPI_Final%20report.pdf (accessed August 22, 2012).
- Spaapen, J., L. van Drooge. 2011. Introducing “productive interactions” in social impact assessment. *Research Evaluation* 20/3: 211–218.
- Stockmann, R. 2004. *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren*. CEval Arbeitspapiere 9. Saarbrücken: Centrum für Evaluation (CEval). www.ceval.de/typo3/fileadmin/user_upload/PDFs/workpaper9.pdf (accessed May 7, 2013).
- Stokols, D. et al. 2003. Evaluating transdisciplinary science. *Nicotine and Tobacco Research* 5/1: 21–39.
- Stoll-Kleemann, S. 2007. Potentiale der Evaluation inter- und transdisziplinärer Nachhaltigkeitsforschung und Humanökologie. In: *Evaluation inter- und transdisziplinärer Forschung. Humanökologie und Nachhaltigkeitsforschung auf dem Prüfstand*. Edited by S. Stoll-Kleemann, C. Pohl. Munich: oekom. 25–40.
- Tåbara, J. D., J. Jaeger. 2012. *Analysis of pathways workshop*. www.visionrd4sd.eu/news/71 (accessed September 20, 2012).
- Uriarte, M., H. A. Ewing, V. T. Eviner, K. C. Weathers. 2007. Constructing a broader and more inclusive value system in science. *BioScience* 57/1: 71–78.
- Van Vught, F., F. Ziegele (Eds.). 2011. *U-Multirank – Design and testing the feasibility of a multidimensional global university ranking: Final report*. CHERPA-Network (Consortium for Higher Education and Research Performance Assessment). http://ec.europa.eu/education/higher-education/doc/multirank_en.pdf (accessed September 12, 2011).
- VDW (Federation of German Scientists). 2010. *Für eine verantwortbare und zukunftsorientierte Forschungspolitik in Deutschland*. VDW Positionspapier. Berlin: VDW. www.vdw-ev.de/images/stories/vdwdokumente/aktuelles/vdw_papier_zur_forschungspolitik.pdf (accessed May 16, 2011).
- Walter, A. I., S. Helgenberger, A. Wiek, R. W. Scholz. 2007. Measuring societal effects of transdisciplinary research projects: Design and application of an evaluation method. *Evaluation and Program Planning* 30/4: 325–338.
- WBGU (German Advisory Council on Global Change). 2011. *World in transition: A social contract for sustainability*. Berlin: WBGU.
- Wolf, B., T. Lindenthal, M. Szerencsits, J. Heß. Forthcoming. *Berücksichtigung der Leistungen für Praxis und Gesellschaft in der Forschungsevaluierung: Praxis-Impact-Synthesebericht*. www.orgprints.org.
- WR (German Council of Science and Humanities). 2013. *Empfehlungen zu einem Kerndatensatz Forschung*. Berlin: WR. www.wissenschaftsrat.de/download/archiv/2855-13.pdf (accessed February 5, 2013).

Submitted January 30, 2013; revised version accepted April 27, 2013.

Birge Wolf



Born 1979 in Bad Gandersheim, Germany. Studies in organic agricultural sciences (Dipl.-Ing. agr.). Since 2009 junior scientist at the University of Kassel, Germany, currently in the project *Development and Testing of a Concept for Documentation and Evaluation of Productive Interactions and Impacts on Practice and Society in Agricultural Research*. Research interests: research evaluation, soil fertility, cropping systems, economics in organic agriculture, urban gardening.

Thomas Lindenthal



Born 1967 in Bregenz, Austria. 2001 PhD in organic agriculture. Since 2009 coordinator at the Centre of Global Change and Sustainability, University of Natural Resources and Life Sciences, Vienna, and project leader at the Research Institute of Organic Agriculture (FiBL) Austria, Vienna. Research interests: sustainability research, CO₂-emissions of foodstuffs, research evaluation.

Manfred Szerencsits



Born 1968 in Güssing, Austria. Studies in environmental engineering, 2007 PhD in organic agricultural sciences. Currently scientist at oeko-cluster.at (intercrops for biogas production) and consultant for agriculture and environmental engineering. Research interests: intercrops for biogas production, groundwater protection, decision processes of persons involved in flood risk management, research evaluation.

J. Britt Holbrook

Born 1970, California, USA. PhD in philosophy from Emory University, 2004. Since 2005 Research Assistant Professor in the Department of Philosophy at the University of North Texas. Since 2008 Assistant Director of the Center for the Study of Interdisciplinarity. Research interests: scholarly communication, research evaluation, impact. ORCID: 0000-0002-5804-0692.

Jürgen Heß



Born 1953 in Wassenberg/Niederrhein, Germany. Professor at the University of Natural Resources and Life Sciences, Vienna, Austria. Since 1997 head of the Department of Organic Farming and Cropping Systems at the University of Kassel, Germany. Research interests: development of crop management systems in organic farming, impact of agricultural research on practice and society. Further areas of expertise: scientific policy advice.