

Interdisciplinarity in research evaluation

Chapter for “*The Oxford Handbook of Interdisciplinarity*”, 2nd edition
Edited by Robert Frodeman, Oxford University Press

Version May 2016

Katri Huutoniemi¹ and Ismael Ràfols^{2,3}

¹ Academy of Finland, Helsinki

² Ingenio (CSIC-UPV), Universitat Politècnica de València, València

³ SPRU (Science Policy Research Unit), University of Sussex, Brighton

Abstract

The evaluation of interdisciplinary research is complicated by ambiguity about what interdisciplinarity is and what it should be. The question is topical, as evaluation plays an important role in how science is being shaped and changed today. The chapter performs a meta-analysis of the concept of interdisciplinarity in research evaluation, and gives an epistemic account of what would be involved in such evaluations. First, we discuss the various ways interdisciplinarity can add value to the disciplinary organization of academia and their respective implications for research evaluation. Second, we provide tools for mapping and measuring these value-added properties and illustrate what kind of evidence they can convey to research evaluations. The combined examination of values and indicators allows us to gain a more differentiated understanding of what exactly to look at when evaluating interdisciplinary research – and more generally, how to design research evaluations from an interdisciplinary point of view.

Introduction

Research evaluation (or quality assessment) means the systematic determination of the merit, worth, and significance of a research activity. It implies the existence of both a judgment of quality and a set of organizational procedures and outcomes that are associated with that judgment (Brennan 2007). The evaluation of interdisciplinary research, however, is a tricky issue. On the one hand, the concept of interdisciplinarity already contains a presupposition that it is a valuable thing, as it offers something that is missed by disciplines. On the other hand, because of its deviation from disciplinarity – one of the conditions of possibility of academic knowledge – its value is somehow dubious. This chapter addresses this dilemma by discussing the expected benefits of interdisciplinarity in knowledge production, and the means and measures by which those benefits may be acknowledged and captured in research evaluation.

Such normative consideration of interdisciplinarity requires a critical awareness of the two principal meanings of the term ‘discipline’: first, it refers to a particular branch of learning or body of knowledge, and second, it refers to the maintenance of order and control (Moran 2002). This echoes the relationship between knowledge and power, which is crucial for considering questions of research evaluation. Evaluation is, in essence, a means of exercising control over knowledge. In the

case of interdisciplinary research, however, there is no consensus on the legitimate sources and types of control over it. Underlying the debate are uncertainties that center on the concept of interdisciplinarity itself, which often comes in different variations, such as multi-, inter-, cross and transdisciplinarity.

First, it is far from clear what defines the quality (or excellence) of interdisciplinary research. Whenever research crosses boundaries between disciplines, the problem arises that each discipline carries specific and sometimes conflicting assumptions about what constitutes quality. The criteria of disciplinary communities are proving insufficient for research that expands, integrates, or challenges the discipline's own canon. In such intellectual exchanges, what exactly is it that determines the relevant criteria: one's own discipline or the other discipline, or some combination of the two – or perhaps knowledge users outside of academia? Second, and related to the previous point, it is unclear who judges interdisciplinary work. Since there is no clearly defined community of peer reviewers as there generally is for disciplinary research, competent reviewers can be very hard to find. Thus, peer review is often biased towards established approaches, unreliable in assessing interdisciplinary work, and helpless in comparing different types of excellence against each other (see Holbrook, this volume). Third, there is no agreement on what constitutes interdisciplinarity, and how it can be identified and measured in practice (Huutoniemi *et al.* 2010). The definitional debate tends to be paralyzed by the notion that interdisciplinary research can have so many profiles (see Klein 2006, 2008).

The present chapter seeks to address these issues by performing a meta-analysis of the concept of interdisciplinarity in research evaluation. In particular, we aim to overcome two major divisions in the existing understanding of the topic. First, there are different normative framings of interdisciplinarity, which shape assumptions about quality and how it is best determined. From this discussion, we identify three major epistemic values or guiding principles of interdisciplinarity and discuss their meaning for research evaluation. Second, there is a growing gap between conceptualizing and measuring interdisciplinarity in research evaluation, and the purposes these endeavors have come to serve. Rather than prioritizing one approach over the other, our aim is to bring them together in a mutually reinforcing way. Parallel to the conceptual discussion of the values of interdisciplinarity, we provide bibliometric approaches for mapping and measuring the cognitive properties of research that can be associated with those values. The combination of qualitative and quantitative definitions makes the chapter particularly useful for the purposes of reconsidering and designing research evaluations from an interdisciplinary point of view. The actual implementation of these definitions is likely to differ between particular evaluative settings and is thus beyond the scope of this chapter.

1 Epistemic values of interdisciplinarity

Evaluations are used, among other things, to certify research activity as valid, to distribute resources in academia, to improve the performance of researchers and organizations, to inform strategic decisions, and to legitimate scientific knowledge in society. Different functions of evaluation raise different questions about interdisciplinarity and offer different kinds of control over knowledge production. In order to better understand and deal with these issues, we review some of the main benefits or 'goods' that interdisciplinarity is expected to convey. We are not so much offering a

procedure of evaluating interdisciplinary research as giving an epistemic account of what would be involved in doing so.

Interdisciplinarity is not an end itself, but a means of advancing knowledge. To this end, it has several assets that are not, or not appropriately, provided by disciplinary research. In scholarly and policy discussions on the epistemic benefits of interdisciplinary research, three overarching values stand out: breadth, integration, and transformation. Following the standard usage of terms (e.g. Klein, this volume), one might classify research pursuing these values as multidisciplinary, interdisciplinary and transdisciplinary, respectively. However, our aim is not to provide specific evaluation criteria for different categories, but to illuminate the various ‘added values’ that may and do span those categories. While the primary focus of this chapter is on interdisciplinarity as an academic endeavor, the values capture the epistemic aspects of research that involves actors beyond the academic realm. In what follows, we discuss the meaning, implications and relevance of these values for research evaluation, and summarize them in Table 1.

1.1 Breadth

The most common value of interdisciplinary research, compared to disciplinary research, is the breadth of subject matter, vision, or skills: its span of attention extends to more than one discipline, field or specialty. The flow of ideas and intellectual exchange across fields are promoted to combat the general tendency of disciplines and specialties to become self-referential, monolithic structures. Such dynamics are often deemed a premise of both intellectual development and problem solving, at least in the long term (e.g. Stirling 2007).

In pursuing breadth, specialized expertise is the baseline for assessment, which is guaranteed by the professional accountability of experts to their respective epistemic communities: ‘A basic premise of quality interdisciplinary work is that it satisfies quality standards arising from the disciplines involved’ (Boix Mansilla *et al.* 2006, p. 73). Keeping disciplinary depth allows scholars to bring their disciplinary specialty to bear in interdisciplinary collaboration. Yet, it places new demands on the organizational arrangements through which an expanded repertoire of expertise is mobilized. The focus of interdisciplinary evaluation is thus on the effective division of cognitive labor across specialties: well-managed coordination, collaboration and exchange are crucial indicators of successful interdisciplinary work.

The central challenge of evaluating breadth is to include an appropriate range of experts in evaluation constituencies, and to handle their likely disparate inputs into the review process. To avoid cognitive particularism and disciplinary parochialism, reviewers should be open-minded, respectful of various traditions, and tolerant to approaches other than their own. The participation of researchers in the selection of appropriate reviewers can ensure that all relevant aspects of the work are taken into account. Deliberation among evaluation constituencies is critical to dealing with different judgments and weighting the different disciplinary contributions.

Here, the stance is to ensure fair and competent evaluation for interdisciplinary research within the existing structures of academia. Yet, fairness is not easily achieved. Several empirical studies have illustrated various biases towards established approaches. Interdisciplinarity is easily considered by peer review panels a ‘plus’, but not substitutive for disciplinary markers of quality (Lamont 2009).

To expect both disciplinary depth and interdisciplinary breadth in a single proposal is, however, to place unrealistic demands on researchers (Lyall *et al.* 2011). Competent evaluation of breadth may be difficult to achieve, too, as collective evaluation processes are often characterized by a clear division of scholarly tasks, little interaction, and tacit compromises (Langfeldt 2004). Interdisciplinary considerations, however, can be encouraged by selecting reviewers with sufficiently overlapping expertise and/or personal interdisciplinary competence (Huutoniemi 2012).

Despite these challenges, the breadth of expertise is a highly relevant criterion for evaluating any large-scale project, program, or organization. It emphasizes the need to manage diversity in the increasingly specialized and complex system of knowledge production. However, it does not help institutionalize interdisciplinary scholarship as a distinctive pursuit in its own right, but builds on the strength, flexibility, and self-organization of existing disciplines. Lack of autonomy and authority over evaluation criteria can make interdisciplinary research unrewarding, risky, and vulnerable to specific disciplinary interests. At the same time, its evaluation will require much bureaucratic effort, which is subject to its own problems (see Jacobs 2013).

Table 1. Three major epistemic values of interdisciplinarity and their implications for research evaluation.

	<i>Breadth</i>	<i>Integration</i>	<i>Transformation</i>
Value added	Expanded repertoire of specialized expertise	Synthesis of perspectives	Transformation of specialized worldviews
Accountability	Multiple disciplines	Integrative research context	Hybrid communities, future generations
Evaluative focus	Management of diversity	Integrative process	Creativity, renewal of knowledge structures
Epistemic standards	Combination of disciplinary standards	Specific standards for integration	Proactive, emergent standards
Policy implications	Structural flexibility in the evaluation process	An evaluation system of its own	New governance of knowledge production
Proponents	Academic organizations, sociologists of science	Problem-oriented organizations, practitioners, and theorists	University reformers, antidisciplinary movements
Pathologies	Increase of bureaucracy, lack of community	Institutional isomorphism with disciplines, including their limitations	Epistemic anarchy, no cumulative advancement

1.2 Integration

Another central value of interdisciplinary research is the capacity to bring together knowledge from disparate fields into a synthetic or coherent whole. This highlights a dimension of interdisciplinarity that is presumed but not problematized in the pursuit of breadth. Synthesis or integration is typically regarded as the distinguishing but elusive characteristic of interdisciplinarity (Repko 2012). Despite the huge variety of interdisciplinary activities, the common bond is often the need to develop an integrated end result, either an intellectual synthesis or a solution to a practical problem (NAS 2005, p. 2). Integration is understood as a means toward greater insight and greater success at problem solving.

The instrumental role of integration in pursuing relevant ends marks a clear departure from disciplinary standards of quality, and provides a point of reference for evaluating the merits of interdisciplinary research in their own right. Underlying this view is the observation that integration is indeed a very complex effort requiring specific concepts, tools and expertise that cannot be reduced to its disciplinary components. Moreover, it is often critical to integrate knowledge from the field of practical action that the research is related to (e.g., Pohl *et al.* this volume).

While the conventional standard of scholarship rests on the mastery of an intellectual domain, interdisciplinary scholarship rests partly on procedural expertise. As integration is a social and cognitive process, a valid assessment must involve some indication of the degree or extent of knowledge integration that occurs while research is being conducted (Wagner *et al.* 2011). This may require assessment during the research alongside the traditional *ex ante* and *ex post* evaluations. Important procedural markers of successful interdisciplinary integration play out in the collectively constructed 'cognitive-emotional-interactive platforms' which operate in conjunction with specific institutional conditions created by funders (Boix Mansilla *et al.* forthcoming).

The pursuit of integration has generated guiding questions for the evaluation of interdisciplinary projects and programs: Is the diversity of disciplines and fields too narrow or too broad for the task at hand? Have relevant approaches, tools, and partners been identified? Has synthesis unfolded through patterning and testing the relatedness of materials, ideas, and methods? Have known integrative techniques been utilized? Is there a unifying principle, theory, or set of questions that provides coherence? (Klein 2008). As the questions indicate, the evaluation of integration typically takes place 'in context': it aims to encompass the various activities of the research group and allow for the influence of relevant stakeholders in the evaluation process (Spaapen *et al.* 2007).

As a new, integrative mode of scholarship, interdisciplinarity calls for an evaluation system of its own (see Stokols *et al.* this volume). To make any system of evaluation work, there needs to be a community of practice with shared norms, values, experiences and referent points. An exemplary description of such a system is Julie Klein's *Creating Interdisciplinary Campus Cultures* (2010). However, sophisticated criteria for interdisciplinary integration do not solve the problem of how to evaluate new syntheses vis-à-vis more discipline-based accounts of the same phenomena, or for that matter, whether and how to incorporate interdisciplinary values into academia as a whole. There are also dangers in ranking interdisciplinarity in terms of degree to which knowledge from disparate fields is brought together in a synthetic or integrative manner. First, integrated solutions from one point of view are often clearly limited or incomplete from another point of view (Jacobs

2013). Second, integration is by no means the goal of all interdisciplinary work, which may be exploratory or critical in intent (Barry and Born 2013), whereas important intellectual syntheses can occur also within disciplines. Third, the very idea of integration neglects the possibility that knowledge created in different conceptual spaces is incommensurable (Holbrook 2013).

1.3 Transformation

The third epistemic value of interdisciplinarity is its potential to transcend or transform the old divisions, disciplines, and dogmas of knowledge (e.g., Barry and Born 2013; Klein 2014). The impetus is that the status quo is not sustainable, as disciplines have failed to understand the pressing challenges of humanity. This stance highlights the fact that knowledge is not separate from politics and action, but influences and is influenced by them. Interdisciplinarity is promoted as a liberating force that challenges the existing structures of knowledge and transcends narrow disciplinary worldviews. It may be associated with critical or emancipatory goals of knowledge, and/or seen as a source of radical innovation and breakthrough.

We call this value ‘transformation’, and detect it behind a heterogeneous set of interdisciplinary activities within and beyond the academia. In the past few decades, transformative interdisciplinarity has been extensively justified by political and societal demands (Gibbons and Nowotny 2001). The contextualization of problems in various real-world settings and the public accountability of science are often set directly against the disciplinary model of knowledge production. However, transformation may also result from an initially apolitical attempt to resolve paradoxes, for example, between different epistemologies. Interdisciplinarity can thus facilitate the process through which science transforms its own institutional design and social relations, expanding its problem space over time.

In any case, implications for evaluation are profound. One possibility is moving away from a closed system of quality control toward an open-ended process that is not held captive by the status quo of existing epistemic categories and their constituencies. The worth of interdisciplinary efforts lies not in their consistency with disciplinary or other institutional antecedents but precisely in their capacity to transform them. While given epistemic standards for such efforts do not exist, priority is placed on positively reframing and refocusing how important real world challenges are addressed. A crucial point is that ‘the design, implementation and interpretation of the entire research . . . is conducted as an equal collaborative partnership with disparate interests *beyond the practitioners themselves*’ (Stirling *et al.* 2015, p. 32, emphasis added). This definition emphasizes the benefits of broadening out and opening up existing understandings.

The evaluation of interdisciplinary transformation requires a proactive stance, including openness to inputs from relevant stakeholders. The inclusion of various stakeholder groups in evaluation is deemed important not so much for ensuring competent evaluation, but for making the epistemic stakes and blind spots visible and open to negotiation. Thus, public engagement is aimed at ‘giving voice’ to those who may have other questions to be answered, other ways of answering them, and other conclusions to draw than one’s disciplinary peers (Spaapen *et al.* 2007). Accordingly, the notion of ‘peer’ is being extended to include all those who have a desire to participate in the resolution of the issue at stake (Funtowicz and Ravetz 1993). Various models of deliberative

democracy are used to foster this ideal, such as ‘consensus conferences’ or ‘citizen juries’ (McDonald *et al.* 2009).

Successfully pursuing and evaluating transformation is confronted with persistent problems regarding the scope of interests and values that are taken into account, and the methods that can be applied to reflexively weigh the evidence of quality. It is not clear that simply extending the stakes involved will make the evaluation process more open to transformative interdisciplinarity. Also the opposite can be the case: There is a danger of politicizing evaluation to the extent that shared epistemic values are overridden by more partisan interests. This, in turn, threatens the internal capacity of scientific inquiry to transform the existing social reality by ‘speaking truth to power’. Moreover, while interdisciplinary transformation is often highly valuable in specific occasions, it builds on and complements to the operation of disciplinary science.

2 Mapping and measures of interdisciplinarity

As a result of the contested nature of interdisciplinarity, qualitative approaches to its evaluation are sometimes dismissed as lacking rigor or being potentially influenced by partisan interests in what constitutes interdisciplinarity and what constitutes scientific quality (Laudel and Origgi 2006). Quantitative approaches can be helpful in documenting and expanding empirical evidence as well as in providing contrasting perspectives on interdisciplinarity (Rafols *et al.* 2012). Such evidence may be important in policy dynamics since quantitative approaches are generally seen as more ‘objective’. However, as we will see, mapping and measuring interdisciplinarity depends on very specific choices on classifications and metrics that are value laden even if based on objective evidence. Yet, combined with deliberation on appropriate values, these quantitative tools can enrich the evaluation of interdisciplinary research, for example in terms of transparency.

This section proposes methods for mapping and measuring interdisciplinarity in terms of breadth, integration, and transformation, in order to illustrate the values presented above. We discuss how each of these concepts can be operationalized in various ways depending on the aspects that are considered relevant for a specific evaluation. In addition to indicators, we also propose visualization tools (*viz.*, science maps) that convey a more nuanced understanding than one-dimensional measures of interdisciplinarity (Rafols *et al.* 2010). The goal of the scientometric analysis is to provide empirical support to the evaluation of breadth, integration and transformation in a given body of research, which constitutes the ‘system’ under evaluation (e.g., a project, a university, a large collaboration, a funding program).

Visualizations and measures of interdisciplinarity require the use of knowledge classifications, which are quite controversial. One may conceive different perspectives for classifying science: not only disciplinary categories, but also, for example, diseases or patent classes. Although user-oriented classifications may play an important role in the evaluation of research impacts, they are not yet robustly developed. In this chapter we will focus on disciplinary categories that do not capture non-academic aspects of research.

Underlying a knowledge classification is the notion of ‘cognitive distance’ (Stirling 2007): some categories can be conceived as more proximate or further apart than others. For example, cell biology and biochemistry are understood as more similar (*i.e.*, more proximate in a cognitive space)

than cell biology and geophysics. Cognitive distance can be operationalized in terms of the correlations (e.g., Pearson or cosine similarity) of some variables (e.g., citations, word occurrence) of disciplinary categories. We use the convention that the higher the citation correlation among knowledge categories, the more conventional the relation between the categories and thus the closer the cognitive distance.

The sections below explain how measures of interdisciplinarity can be applied for a given body of research that is represented by a set of publications (or other documents) produced by a given organization (a laboratory, a department, a university), a small or large project, or a research area.

2.1 Operationalization of breadth and integration

We propose to operationalize ‘breadth’ with the concept of diversity (Stirling 2007) and ‘integration’ with a combination of diversity and coherence (Rafols *et al.* 2012). Diversity aims to capture the distribution of elements (e.g., researchers, publications, financial resources) across categories (e.g., disciplines) for a given body of research. Coherence describes the extent to which elements of different categories are related, for example via interactions between researchers, via exchanges of information such as letters or e-mails, or via citations. Figure 1 illustrates the two concepts. ‘Integration’ is characterised both by high diversity *and* high coherence because it implies that linkages are made between somewhat disparate elements.

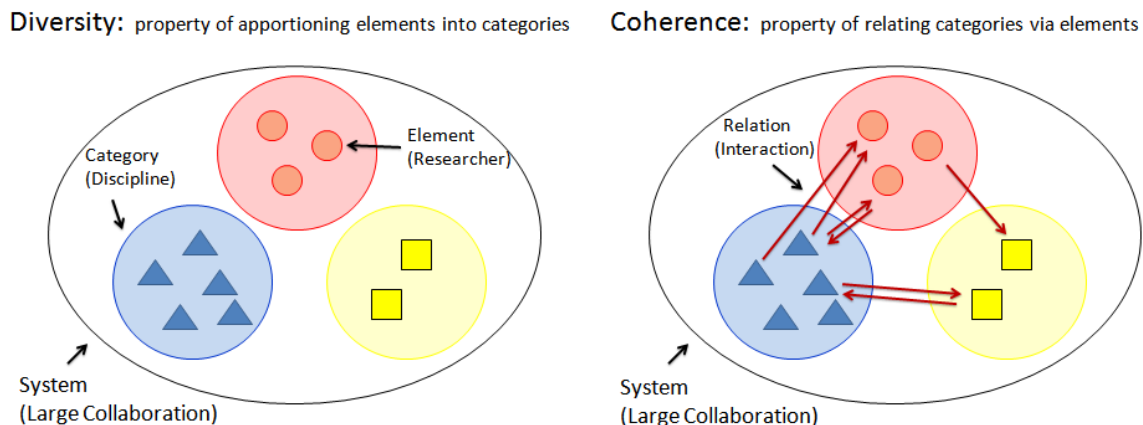


Figure 1. Illustration of diversity (left) and coherence (right). Large circles represent categories. Small figures represent elements of a specific category (triangles, circles, squares). In parentheses, an example of operationalizing the concepts in the evaluation of a large collaboration: diversity refers to the allocation of researchers (elements) across disciplines (categories), and coherence refers to the cross-disciplinary interactions among researchers. Based on Rafols (2014).

The mathematical operationalizations of diversity and coherence are not unique. As illustrated in the left side of Figure 2, diversity can increase when the number of categories increases (variety), when the distribution of elements across categories become more even (balance), or when the elements are distributed across more distinct categories (disparity) (Stirling 2007). Also, as illustrated in the right side of Figure 2, coherence can increase when the number of relations increases (density), when the relations become stronger (intensity), or when they link more different categories (disparity) (Rafols 2014).

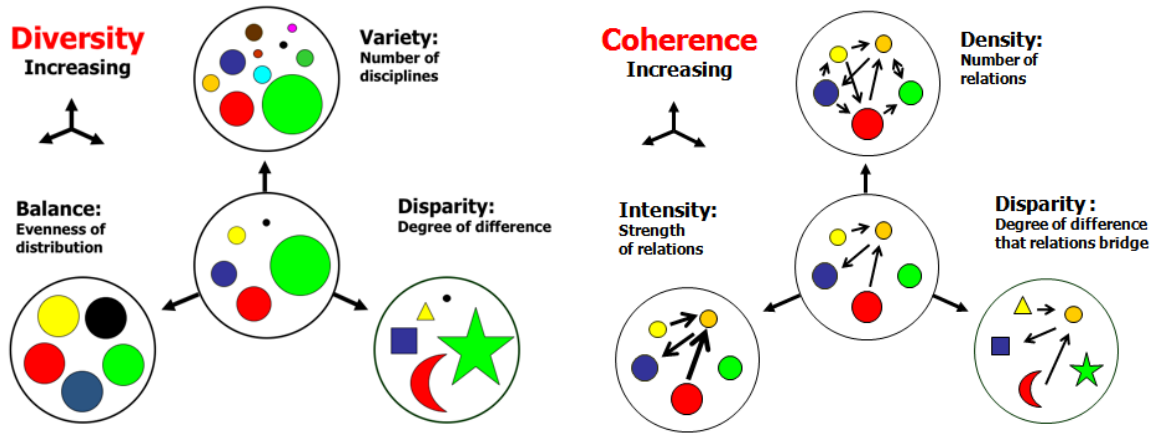


Figure 2. Schematic representation of the attributes of diversity (left) and coherence (right). Each full circle represents a system under study. The figures inside the circle are the categories into which the elements are apportioned. Different shapes indicate more difference between the categories. The size of the figures indicates the proportion of elements in a category. Thicker lines indicate higher intensity in relations. Based on Stirling (1998) and Rafols (2014).

Table 2. Measures of diversity and coherence. The measures which have been used and tested in the literature are highlighted.

Notation:	
Proportion of elements in category i :	p_i
Intensity of relations between categories i and j :	i_{ij}
Cognitive distance between categories i and j :	d_{ij}
Diversity Indices:	
Generalised Stirling diversity	$\sum_{i,j(i \neq j)} (p_i p_j)^\alpha d_{ij}^\beta$
Variety ($\alpha=0, \beta=0$)	N
Simpson diversity ($\alpha=1, \beta=0$)	$\sum_{i,j(i \neq j)} p_i p_j$
Rao-Stirling diversity ($\alpha=1, \beta=1$)	$\sum_{i,j(i \neq j)} p_i p_j d_{ij}$
Coherence Indices:	
Generalised Coherence	$\sum_{i,j(i \neq j)} i_{ij}^\gamma d_{ij}^\delta$
Density ($\gamma=0, \delta=0$)	M
Intensity ($\gamma=1, \delta=0$)	$\sum_{i,j(i \neq j)} i_{ij}$
Coherence ($\gamma=1, \delta=1$)	$\sum_{i,j(i \neq j)} i_{ij} d_{ij}$

Any single index of diversity captures the three different properties of variety, balance, and disparity but may weight them differently. Likewise, any measure of coherence makes an implicit choice of the relative weight of density, intensity, and disparity. Following Stirling (2007), we have proposed generalized heuristics and measures for exploring diversity and coherence as shown in Table 2 and discussed in Rafols (2014).

Breadth and integration are operationalized here at a high level of abstraction in order to allow for using various data sources and fit to different contexts. Bibliometric data can be used to calculate the various measures, including cognitive distance, which is difficult to estimate without structured data (Rafols *et al.* 2012).

Visualizations can be very helpful in conveying information on diversity and coherence without complex mathematics (see Box 1). This is illustrated in Figure 3, which shows the citations made by the publications of a research institute (ISSTI, University of Edinburgh). This information illustrates areas in which ISSTI is active, and how ISSTI is unique in linking disparate disciplines in comparison with conventional patterns. Evidence of such linkages enables, for example, deliberations about the success of a project in making specific connections between e.g. management and energy research.

How science maps can inform evaluations

Figure 3 illustrates the disciplinary diversity (breadth) and coherence (integration) of an institute (ISSTI, *The Institute for the Study of Science, Technology and Innovation*, University of Edinburgh).

It captures the three properties of diversity: (1) the number of circles reveals the variety of disciplinary categories; (2) the relative size of the circles shows whether the citations are concentrated in a few disciplines or are more evenly spread (balance); and (3) the distance between the circles tells whether the publications spread over very different research areas (disparity of distribution).

Also, the viewer can intuitively become acquainted with the notion of coherence: (1) the number of lines illustrates the density of citations; (2) the thickness of lines reveals the intensity of citations across disciplinary categories; and (3) the distance crossed by the lines shows the disparity of links – i.e. if citations are made across fields that do not often interact and therefore lie far apart in the map of science.

As shown in Figure 3, one can compare the observed citation patterns with the expected citation patterns for a specific publication set, and see if the observed patterns stand out as more dense, more intense, or more disparate than the expected patterns. Thus, the

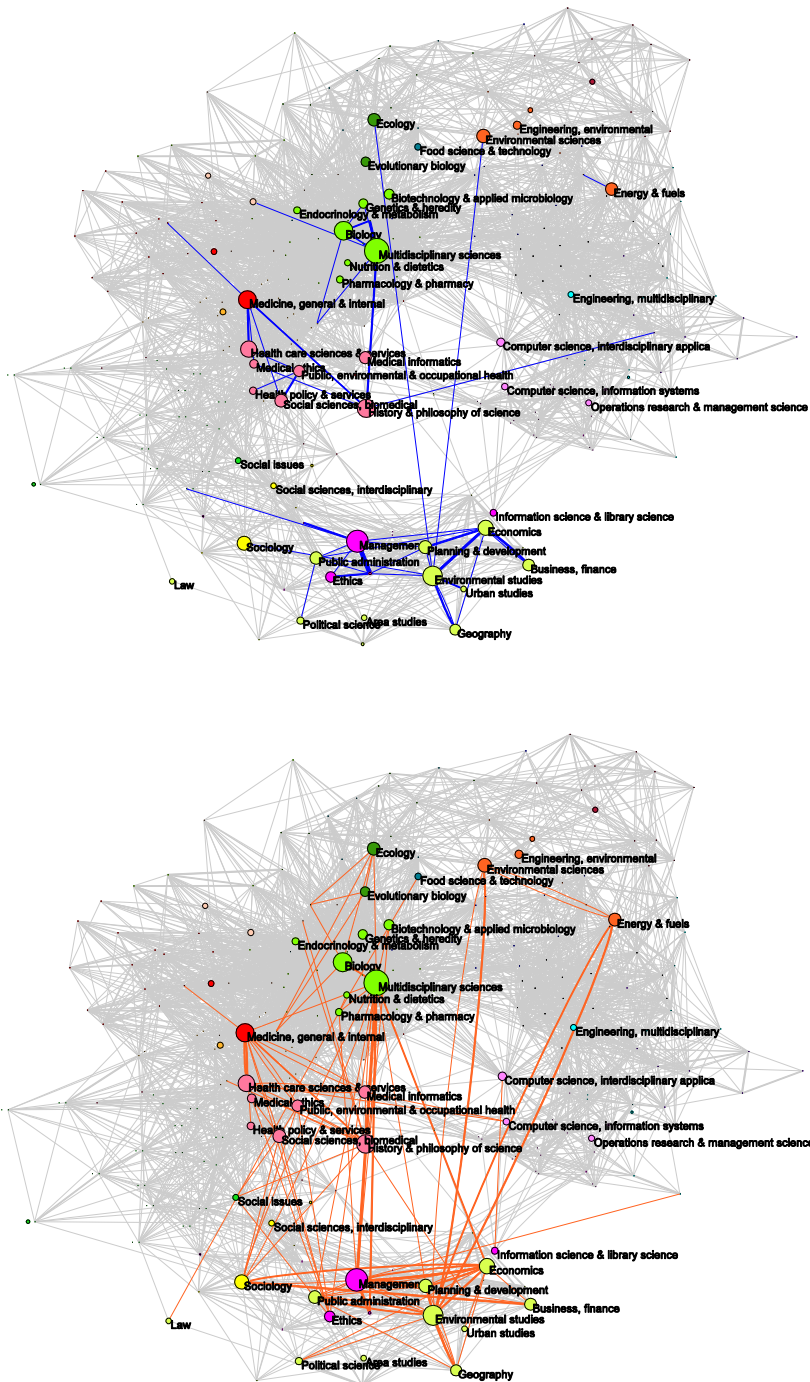


Figure 3. Expected (top) and observed (bottom) citations of the research centre ISSTI (University of Edinburgh) across different Web of Science categories. The grey lines in the background position disciplinary categories in the global map of science (Rafols et al. 2010). The size of the nodes illustrates the aggregate number of citations given to a category from all ISSTI's publications (2006-2010). Blue lines (top) show the expected citations between the specific categories in which ISSTI publishes. The computation of expected citations is based on the total number of publications in a category, and the average proportion of citations to all other fields. It can be observed that the expected citations tend to be within disciplines: within biological sciences, within health services, and within social sciences. Orange lines (bottom) show the citations between fields observed in ISSTI's publications. Source: Rafols et al. (2012)

2.2 Operationalization of transformation

Transformation is more difficult to map and measure than breadth and integration, because, by definition, it cannot be captured by pre-existing disciplinary categories. Conventional disciplinary categories are based on institutionalized classification systems which are slow to capture changes in the knowledge landscape. Using old disciplinary categories over new knowledge may create measurement artifacts. It is thus necessary to create descriptions of the scientific landscape that are sufficiently fine-grained to capture developments that fill up previously empty cognitive spaces or change the overall knowledge structure. Thus, it becomes necessary to use units of analysis such as research specialties or research topics rather than disciplines. Journals can be used for some purposes, but they are sometimes problematic because many important journals cover several specialties and topics. Co-word maps, topic modeling or small topic clusters may provide a richer, more fine-grained and more reliable base for knowledge landscape to capture transformations.

One way to spot the transformative nature of an interdisciplinary effort is to check whether it falls outside the main disciplinary concentrations. This is illustrated in Figure 4 for the case of Management and Innovation Studies. The journal maps show areas of high density that correspond to the disciplinary cores of Management and Economics. The areas of low density zone are journals related to Science, Technology, and Innovation, which are less dominant and can represent a transformation of the landscape. Ideally, one would look at these maps over time to see how the landscape is developing. But with a static map such as this, one may hypothesize that publications in the interstitial or peripheral areas are more related to efforts to transform science than those in the disciplinary cores. The overlay maps of ISSTI and London Business School (LBS) show contrasting patterns of publication: ISSTI publishes in more transformative areas, whereas LBS publishes mainly in the disciplinary cores. This information can be useful in evaluations that aim to trace interdisciplinary transformation.

3 Conclusion

As suggested at the outset, the evaluation of interdisciplinary research is complicated by ambiguity about what interdisciplinarity is and what it should be. We have addressed this problem by analyzing the major expectations of interdisciplinarity in both qualitative and quantitative terms. First, we have discussed the various ways interdisciplinarity can add value to the disciplinary organization of academia and their respective implications for research evaluation. Second, we have provided tools for mapping and measuring these value-added properties and illustrated what kind of evidence they can convey to evaluation.

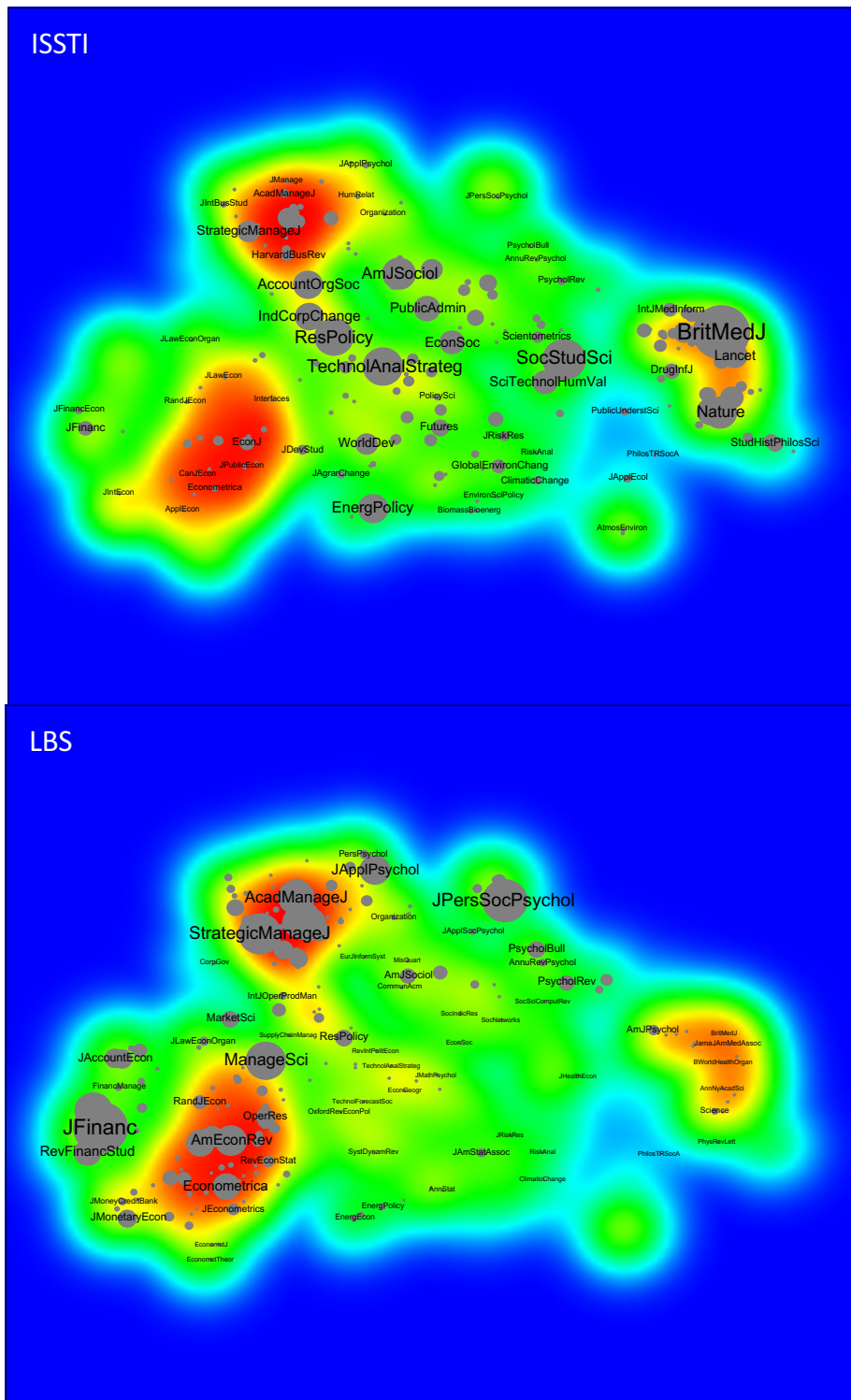


Figure 4. Overlay of the references (grey nodes) of the research centre ISSTI (University of Edinburgh) and London Business School (LBS) publications over a journal map. The map illustrates the similarity structure of the 391 most important journals in Management and Innovation studies. Red areas correspond to a high density of journals, indicating areas of conventional disciplinary activity. Green areas show low density. The size of grey nodes indicates the proportion of the research unit's references in a given journal. Journals located in between red areas, i.e. between disciplinary cores, are interpreted as potential areas of transformation. Source: Rafols et al. (2012).

The combined examination of values and indicators allows us to gain a more differentiated understanding of what exactly to look at when evaluating interdisciplinary research. The first step of the evaluation of interdisciplinary research, we suggest, is to consider the relevance of the various values that interdisciplinary interaction involves (e.g., breadth, integration, transformation). The second step is to select the categories of knowledge (e.g., disciplines, research specialties, technology classes) that can be used as reference points in detecting those interactions. The third step is to select the unit(s) of analysis (e.g., researchers, publications, financial resources) that represents the knowledge base of the given entity (e.g., a university, a network, a funding program). The final step is to enquire into the degree and form of interdisciplinarity in terms of the diversity, coherence and/or transformation of the selected units.

In the literature, it is often acknowledged that interdisciplinarity is not driven by a single goal, and that the variability of goals, in turn, drives variability of criteria and indicators of quality (Klein 2008). In this chapter, we have aimed at a more systemic view of the benefits and indicators of interdisciplinarity. The central values of interdisciplinarity are not exclusive to interdisciplinary activities only, but clearly resonate with the overall goals of science. At the same time, interdisciplinary research is in a good position to advance these goals.

The degree of emphasis placed on each value depends on the purpose of a given evaluation. Measures and maps are only supportive tools in order to trace the values. Research evaluation is used for so many different purposes and in so many different scopes, levels, and contexts that we have deliberately not addressed such issues here. Beyond any particular perspective, however, evaluation is worthy of attention because it is an important part of the way in which science is being shaped and changed today. The incorporation of interdisciplinary concerns in research evaluation is one of the most significant dynamics of such change. An implication of this dynamics is increasing awareness of disciplinary discrepancies, ambiguities, and ignorance, pointing to the need to go beyond disciplinary criteria of validating knowledge. Interdisciplinary considerations in research evaluation, therefore, are relevant and consequential for disciplinary research, too, and should become a routine part of quality control in science (Huutoniemi 2015). At the same time, criteria for interdisciplinary research need to be subjected to critical examination in terms of their systemic effects beyond the particular purpose they are designed to serve.

More explicit discussion of the various purposes and diverse beneficiaries of interdisciplinary evaluation, especially vis-à-vis disciplinary evaluation, is needed for making robust decisions on which values count in specific situations and how their realization can be measured. As we have seen, breadth, integration, and transformation are not equally relevant criteria for all purposes of evaluation, but highlight different, though not incompatible, normative goals. Similarly, the selection of quantitative tools to gauge interdisciplinary properties depends on a number of decisions that are both value-laden and significant for the kind of interdisciplinary relationships that are recognized.

Acknowledgements

KH acknowledges support from Kone Foundation. IR acknowledges support from an EC Marie Curie Reintegration Grant. The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the funders.

References

- Barry, A. and Born, G. (eds.) (2013). *Interdisciplinarity: reconfigurations of the social and natural sciences*. London: Routledge.
- Boix Mansilla, V., Feller, I. and Gardner, H. (2006). Quality assessment in interdisciplinary research and education. *Research Evaluation* 15(1), 69–74.
- Boix Mansilla, V., Lamont, M. and Sato, K. (forthcoming). Successful interdisciplinary collaborations: the contributions of shared socio-emotional-cognitive platforms to interdisciplinary synthesis. *Science, Technology & Human Values*.
- Brennan, J. (2007). The multiple functions of evaluation and quality assessment. In P. Cavalli, ed. *Quality assessment for higher education in Europe: problems, practices and solutions*, pp. 17–26. London: Portland Press.
- Funtowicz, S. O. and Ravetz, J. R. (1993). Science for the post-normal age. *Futures* 25(7), 739–55.
- Gibbons, M. and Nowotny, H. (2001). The potential of transdisciplinarity. In J.T. Klein et al., eds. *Transdisciplinarity: joint problem solving among science, technology, and society*, pp. 67–80. Basel: Birkhäuser
- Holbrook, J. B. (2013) What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese* 190(11), 1865-1879.
- Huutoniemi, K. (2012). Communicating and compromising on disciplinary expertise in the peer review of research proposals. *Social Studies of Science* 42(6), 900–924.
- Huutoniemi, K. (2015). Interdisciplinarity as academic accountability: prospects for quality control across disciplinary boundaries. *Social Epistemology*, DOI: 10.1080/02691728.2015.1015061
- Huutoniemi, K., Klein, J.T., Bruun, H. and Hukkinen, J. (2010). Analyzing interdisciplinarity: typology and indicators. *Research Policy* 39(1), 79–88.
- Jacobs, J. (2013). *In defense of disciplines: interdisciplinarity and specialization in the research university*. Chicago: The University of Chicago Press.
- Klein, J. T. (2006). Afterword: the emergent literature on interdisciplinary and transdisciplinary research evaluation. *Research Evaluation* 15(1), 75–80.
- Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: a literature review. *American Journal of Preventive Medicine* 35, S116–S123.
- Klein, J. T. (2010). *Creating interdisciplinary campus cultures: a model for strength and sustainability*. San Francisco: Jossey-Bass.
- Klein, J. T. (2014). Discourses of transdisciplinarity: looking back to the future. *Futures* 63C, 68–74.
- Lamont, M. (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.
- Langfeldt, L. (2004). Expert panels evaluating research: decision-making and sources of bias. *Research Evaluation* 13(1), 52–62.
- Laudel G., Origg, G. (2006). Introduction to a special issue on the assessment of interdisciplinary research. *Research Evaluation* 15(1), 2–4.
- Lyall, C., Tait, J., Meagner, L., Bruce, A. and Marsden, W. (2011). Short guide to evaluating interdisciplinary research. *ISSTI Briefing Note* (Number 9) March 2011. Available at http://www.ei.udelar.edu.uy/resources/1/9/0/1/0_0d69f2c521bb166/19010_c3bef21aec0471d.pdf
- McDonald, D., Bammer, G and Deane, P. (2009). *Research integration using dialogue methods*. Canberra: Australian National University E Press.
- Moran, J. (2002). *Interdisciplinarity: the new critical idiom*. London: Routledge.

- NAS (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press.
- Rafols, I., Porter, A. L. and Leydesdorff, L. (2010). Science overlay maps: a new tool for research policy and library management. *Journal of the American Society for Information Science and Technology* 61(9), 1871–1887.
- Rafols, I., Leydesdorff, L., O’Hare, A., Nightingale, P. and Stirling, A. (2012). How journal rankings can suppress interdisciplinarity: the case of innovation studies and business and management. *Research Policy* 41(7), 1262–1282.
- Rafols, I. (2014). Knowledge integration and diffusion: measures and mapping of diversity and coherence. In Y. Ding, R. Rousseau and D. Wolfram, eds. *Measuring scholarly impact*, pp. 169–190. New York: Springer. Available at <http://arxiv.org/abs/1412.6683>
- Repko, A. (2012). *Interdisciplinary research: process and theory*. 2nd edn. Los Angeles: Sage.
- Spaapen, J., Dijstelbloem, H. and Wamelink, F. (2007). *Evaluating research in context: a method for comprehensive assessment*, 2nd edn. Hague: Consultative Committee of Sector Councils for Research and Development (COS).
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface* 4(15), 707–719.
- Stirling, A., Nexus Network Team, SPRU and STEPS Centre (2015). *Developing ‘nexus capabilities’: towards transdisciplinary methodologies*. Available at <http://www.thenexusnetwork.org/wp-content/uploads/2015/09/Transdisciplinary-Nexus-Methods-Workshop-final-Report-2015.pdf>
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I. and Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. *Journal of Informetrics*, 5(1), 14–26.