Journal of Informetrics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature

Caroline S. Wagner[a,*], J. David Roessner[a], Kamau Bobb[a], Julie Thompson Klein[b], Kevin W. Boyack[c], Joann Keyton[d], Ismael Rafols[e], Katy Börner[f]

[a] *SRI International, Arlington, VA, USA*
[b] *Wayne State University, Detroit, MI, USA*
[c] *SciTech Strategies, Inc., Albuquerque, NM, USA*
[d] *North Carolina State University, Raleigh, NC, USA*
[e] *Science Policy Research Unit, University of Sussex, Brighton, UK*
[f] *Indiana University, Bloomington, IN, USA*

ARTICLE INFO

Article history:
Received 31 July 2009
Received in revised form 10 June 2010
Accepted 14 June 2010

Keywords:
Interdisciplinary
Science
Research
Indicators
Bibliometrics
Evaluation

ABSTRACT

Interdisciplinary scientific research (IDR) extends and challenges the study of science on a number of fronts, including creating output science and engineering (S&E) indicators. This literature review began with a narrow search for quantitative measures of the output of IDR that could contribute to indicators, but the authors expanded the scope of the review as it became clear that differing definitions, assessment tools, evaluation processes, and measures all shed light on different aspects of IDR. Key among these broader aspects is (a) the importance of incorporating the concept of knowledge integration, and (b) recognizing that integration can occur within a single mind as well as among a team. Existing output measures alone cannot adequately capture this process. Among the quantitative measures considered, bibliometrics (co-authorships, co-inventors, collaborations, references, citations and co-citations) are the most developed, but leave considerable gaps in understanding of the social dynamics that lead to knowledge integration. Emerging measures in network dynamics (particularly betweenness centrality and diversity), and entropy are promising as indicators, but their use requires sophisticated interpretations. Combinations of quantitative measures and qualitative assessments being applied within evaluation studies appear to reveal IDR processes but carry burdens of expense, intrusion, and lack of reproducibility year-upon-year. This review is a first step toward providing a more holistic view of measuring IDR, although research and development is needed before metrics can adequately reflect the actual phenomenon of IDR.

1. Purpose of this literature review

Increases in interdisciplinary research (IDR) have prompted a number of reports and an expanding literature on the performance measures, management, and evaluation of IDR. This literature review began as a response to a request from the U.S. National Science Foundation to identify quantitative output measures (Wagner, Roessner & Bobb, 2009). Deliberations led us to expand the inquiry beyond quantitative measures to be inclusive along the following lines:

* Corresponding author at: School of International Affairs, Pennsylvania State University, University Park, PA 16802, USA. Tel.: +1 814 865 4294; fax: +1 814 867 0405.
E-mail address: csw16@psu.edu (C.S. Wagner).

1751-1577/$ – see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.joi.2010.06.004

1. Measurement of interdisciplinary research should recognize and incorporate input (consumption) and process value (creation) components as well as output (production) while factoring in short-, middle-, and long-term impacts such as delayed citation or patent citation.[1]
2. Interdisciplinary research involves both social and cognitive phenomena, and both these phenomena should be reflected in any measure or assessment of interdisciplinarity.[2]
3. Assessment of research outputs should be broadened beyond those based in bibliometrics, while also factoring in differences in granularity and dimensions of measurement and assessment.[3]

This article explores ideas from each of the related literatures mentioned above, and more particularly, the current and potential interaction space between them. This review also explores several promising methods for creating indicators for policymakers, research managers, evaluators, and students of the research enterprise.

2. Defining terms

The literature has not settled upon common terms to describe the underlying phenomena. A cluster of related terms describes a fairly wide range of phenomena including social and cognitive processes. The IDR literature assumes an underlying disciplinary structure, although few articles on interdisciplinarity begin by defining *discipline* or *field*. (We follow Porter, Roessner, Cohen, and Perreault (2006) who cite the work of Darden and Maull (1977), defining a discipline of science as having a central problem with items considered to be facts relevant to that problem, and having explanations, goals, and theories related to the problem.) The disciplinary structure of science is an artifact of nineteenth and twentieth century social and political organization. Klein (1996) notes that the modern concept of a scientific 'discipline' has been in common use for about a century. Disciplines evolved slowly in the eighteenth century, and then more quickly in the nineteenth century as dedicated funds began supporting professional laboratories such as those of Pasteur and the Curies. The academic form of the larger process of specialization in labor, and the restructuring of higher education resulting in transmission of the German model of the research university to the United States also accelerated the development of disciplines.

Disciplinary distinctions grew more robust and their members more isolated from one another as the academy grew in the nineteenth and twentieth centuries. By the mid-twentieth century, scientific observers such as Boulding (1956) bemoaned the isolation of the scientific disciplines. In the early 1960s, when Derek de Solla Price documented the exponential growth in the number of scientists and scientific abstracts, distinctions among disciplines were the unquestioned norm (Price, 1963). Nevertheless, the mid-century isolation of disciplinary silos gave way quickly to boundary crossing. According to Klein (2008b), the most widely used schema for defining IDR (i.e., multidisciplinary, interdisciplinary, transdisciplinary) derives from a typology presented at the first international conference on interdisciplinary research and teaching in 1970. Schmidt (2008) suggests that IDR arose with a trend towards problem orientation in science and engineering research: 'Interdisciplinarity is viewed as a highly valued tool in order to restore the unity of sciences or to solve societal-pressing problems. Normative aspects are always involved' (p. 58). The Hybrid Vigor Institute report (Rhoten, Caruso, & Parker, 2003) and the National Academies report on facilitating interdisciplinary research (2005) also cite the character of problems as one reason for the rise in interdisciplinarity. This theme – one of emphasizing the application orientation influencing the emergence of interdisciplinary research – is also noted by Van den Besselaar and Heimeriks (2001).

In a 1978 book on science indicators, Garfield, Malin, and Small (1978) discuss interdisciplinarity as 'linkages between specialties of diverse subject matter' (p. 189). Stokols, Hall, Taylor, and Moser (2008) special edition of the *American Journal of Preventive Medicine* (Vol. 35(2), Suppl. 1) on the science of team science includes terms such as unidisciplinarity, multidisciplinarity, interdisciplinarity, and transdisciplinarity with definitions also drawn from Rosenfield (1992). For this review, we have refined several of these definitions and adopted a slightly more generalized set of definitions well based in the literature. These definitions appear in Table 1.

The definitions range from assuming literal, physical interaction to a broader cognitive context discussed in the literature such as going beyond a normative focus to encompass a new way of knowing that grows out of shifts in epistemics, institutional structure, and culture, as favored by Klein. Socio-cognitive-based definitions view IDR as resulting from integration in the mind of a single person. Rafols and Meyer (2010) suggest that *interdisciplinary* is not the right term to explain the cognitive dynamics at the boundaries of disciplines. They assert that cognitive diversity in relation to disciplines, specialties, technologies, industries, stakeholders, and research fronts would be a more appropriate label for activities across disciplines. Klein says the term 'interdisciplinary' is notable for conflicting meaning. She views the solitary term as no longer adequate to describe the underlying phenomena (1996).

[1] Incorporating inputs and process value creation may need to be staged temporally because less data are available, and less has been codified in a way that could be scaled for metrics. Additionally, the dynamics of research, such as impact delays, vary widely across fields due to factors such as publishing delays.

[2] There is a concern that conventional wisdom (particularly in policy-making) has often implicitly equated IDR practice with IDR collaboration. Care needs to be taken about the wording, so as not to feed into this confusion. In particular, policies dependent on ex-ante social disciplinary distance may favor those researchers who did not make previous efforts to approach other fields.

[3] Measurement, assessment, and evaluation are challenging processes on their own, even without trying to incorporate them into the IDR discussion.

Table 1
Definitions of key terms used in this review.

Multidisciplinary approaches juxtapose disciplinary/professional perspectives, adding breadth and available knowledge, information, and methods. They speak as separate voices, in encyclopedic alignment, an ad hoc mix, or a mélange. Disciplinary elements retain their original identity. In short, the multidisciplinary research product is no more and no less than the simple sum of its parts.
Interdisciplinary approaches integrate separate disciplinary data, methods, tools, concepts, and theories in order to create a holistic view or common understanding of a complex issue, question, or problem. The critical indicators of interdisciplinarity in research include evidence that the integrative synthesis is different from, and greater than, the sum of its parts:
• Micro-combinations of models or global schemes that unify disparate approaches
• Consulting and partnering modes, not multidisciplinary contracting of services
• Coordinated and collaborative inputs and organizational framework
• Formation of a new community of knowers with a hybrid interlanguage
• Generation of new insights and disciplinary relationships
• A larger, more holistic understanding of the core problem or question
• Altered perspectives and revised hypotheses.
Transdisciplinary approaches are comprehensive frameworks that transcend the narrow scope of disciplinary worldviews through an overarching synthesis, such as general systems, policy sciences, feminism, sustainability, and the notion of 'transdisciplinary science' in cancer research as a form of 'transcendent interdisciplinary research' that fosters systematic theoretical frameworks for defining and analyzing social, economic, political, environmental, and institutional factors in human health and well-being. More recently, the term has also connoted a new mode of knowledge production that draws on expertise from a wider range of organizations, and collaborative partnerships for sustainability that integrate research from different disciplines with the knowledge of stakeholders in society. Here too, the transdisciplinary product is greater than the sum of its parts, though the scope of the overall effort is more comprehensive and the parts may be more diverse.

Sources: Stokols et al. (2003a, 2003b).

The work of the Interdisciplinary Studies Project of the Harvard Graduate School of Education's Project Zero and the National Academy of Sciences suggest that consensus has emerged around at least one part of the phenomena under consideration: the importance to interdisciplinarity of the process of knowledge integration. Miller and Mansilla (2004) outline four modes of increasing integration of bodies of knowledge in groups – here assuming a social process:

- Mutual ignorance of other disciplinary perspectives.
- Stereotyping that may have significant misconceptions about the other's approach.
- Perspective-taking where individuals can play the role of, sympathize with, and anticipate the other's way of thinking.
- Merging of perspectives has been mutually revised to create a new hybrid way of thinking.

Porter, Cohen, Roessner, and Perreault (2007), following the National Academies report (2005), define IDR as requiring an integration of concepts, techniques, and/or data from different fields of established research. This definition does not presume the presence of teaming. Rafols and Meyer (2010) also follow the National Academies definition: 'Thus the process of integrating different bodies of knowledge rather than transgression of disciplinary boundaries per se, has been identified as the key aspect of so-called 'interdisciplinary research" (p. 2).

A focus on knowledge integration raises its own issues with regard to understanding IDR. Integration is at base a cognitive process, whether it takes place within an individual's mind or within a group, so that a valid assessment of the interdisciplinarity of research must involve some indication of the degree or extent of knowledge integration that took place *as the research was being conducted*. Integration also entails negotiation of conflict and achievement of synthesis, as the Miller and Mansilla (2004) modes of increasing integration attest. This enlarged thinking about the underlying processes is clearly reflected in several of the articles in the 2006 special issue of *Research Evaluation* on evaluating interdisciplinary research (Vol. 15, Issue 1), and especially in Klein's (2008a) literature review. Porter and Rafols (2009) define integration as reflecting a diversity of knowledge sources by research teams or by individuals. Rafols and Meyer (2010) also use a concept of integration in their paper that reflects two aspects of knowledge systems: (a) diversity, or the number, balance, and degree of difference between the bodies of knowledge concerned; and (b) coherence, or the extent to which specific topics, concepts, tools, and data used in a research process are related. This definition is similar to the measure suggested by Garfield et al. (1978).

3. Context: processes and outcomes of interdisciplinary research

This section discusses the processes that constitute IDR, and the outputs of research that could be tapped to measure IDR. The process of integration – whether cognitive or social – is more difficult to observe (and measure) than are the results of the process, which are largely found in published literature. This may explain why more literature has focused on the outputs of research rather than the processes. Nevertheless, both the process and the output are discussed in various parts of the literature. The August 2008 special issue of the *American Journal of Preventive Medicine* devoted to the science of team science includes numerous articles that describe the methods, measures, and results of substantial investment in studying large-scale team science programs (Croyle, 2008; Gray, 2008; Hall, Feng, Moser, Stokols, & Taylor, 2008; Hall, Stokols et al., 2008; Kessel & Rosenfeld, 2008; Klein, 2008a; Mâsse et al., 2008; Nash, 2008; Stokols, Misra, Moser, Hall, & Taylor, 2008). As Stokols, Hall et al. (2008) and Stokols, Misra et al. (2008) and claim:

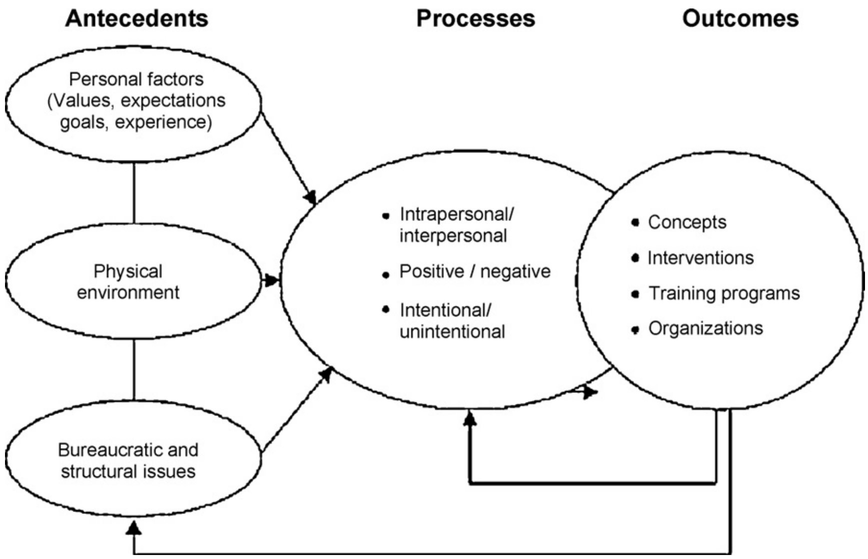

Fig. 1. Working model of transdisciplinary scientific collaboration, from (Stokols et al., 2003a, 2003b).

> The field [science of team science] as a whole focuses not on the phenomena addressed by particular team science initiatives (e.g., cancer, heart disease, obesity, community violence, environmental degradation), but rather on understanding and enhancing the antecedent conditions, collaborative processes, and outcomes associated with team science initiatives more generally, including their scientific discoveries, educational outcomes, and translations of research findings into new clinical practices and public policies. (p. S78)

For the purposes of this review, the distinguishing feature of the emerging field of science of team science is on initiatives of cross-disciplinary teams that 'strive to combine and, in some cases, to integrate concepts, methods, and theories drawn from two or more fields' (p. S78) assuming and not otherwise defining a field or discipline. The introductory article in the August 2008 special issue focuses on interdisciplinary and transdisciplinary science initiatives in which an explicit goal of the collaboration is precisely this integration. The methods and measures discussed have been applied to the antecedents, processes, and outcomes of team science initiatives. The basic conceptual framework used in much of this research is focused on transdisciplinary science, which includes quantitative and qualitative assessment of research processes and outputs (see Fig. 1).

Antecedents include personal factors such as values, expectations, goals, and experience; the physical environment; and bureaucratic and structural issues. Processes include interpersonal, intrapersonal, positive, negative, intentional and unintentional activities. Outcomes include concepts, interventions, training programs, and organizations. Thus the outcome measures used in these studies focus on the effectiveness of IDR programs and processes, presuming their value, rather than on the question of measuring the interdisciplinarity of outputs. Analyses seek to link team composition, management, and contexts to effectiveness measures. Stokols, Hall et al. (2008) and Stokols, Misra et al. (2008) also highlight the multi-methodological approach, noting that 'The combined use of survey, interview, observational, and archival measures in evaluations of team science initiatives affords a more complete understanding of collaborative processes and outcomes than can be gained by adopting a narrower methodological approach' (p. S82). From the point of view of those tasked with creating output indicators, however, these combinations of qualitative and quantitative measures are difficult to reproduce year-on-year and so cannot be considered as useful for creating indicators.

Qualitative measures seek to detect integration in research processes, to assess the value of the outcomes of collaborative work, and to develop causal inferences about the factors that influence highly valued outcomes. Hall, Feng et al. (2008) and Hall, Stokols et al. (2008) report on the year-one evaluation of the National Cancer Institute's Transdisciplinary Research on Energetics and Cancer initiative – a program established as an interdisciplinary project (as opposed to one that emerges as IDR from the nature of the research question). Qualitative measures of interdisciplinary and transdisciplinary collaborations were based on self-assessments by participants using scales of the level of collaboration that occurred. A peer review process, using protocols developed by an evaluation team, was used to assess the degree of cross-disciplinary integration and conceptual breadth of written outputs from the teams.

In the same volume, Mâsse et al. (2008) describe a five-point Likert-type response format used and results obtained from an assessment of processes related to working in transdisciplinary teams. Survey data employing the scales were used by 216 researchers and research staff participating in the NIH Transdisciplinary Tobacco Use Research Centers Initiative. Four scales were developed and tested: three assessed satisfaction, impact, and trust and respect; the fourth assessed transdisciplinary integration. Nash (2008), following Mitrany and Stokols (2005), identifies a mix of qualitative and quantitative indicators that could evaluate transdisciplinary training including hypotheses generated that synthesize theoretical frameworks from different disciplines.

Please cite this article in press as: Wagner, C. S., et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.06.004

Finally, Klein (2008a) places IDR into an evaluative framework that includes seven generic principles including consideration of the variability of goals, criteria, indicators, the levels of integration (social and cognitive organization), the special requirements of leadership, and the unpredictability of long-term impact. She emphasizes that 'Appropriate evaluation. . . . evolves through a dialogue of conventional and expanded indicators of quality. Traditional methodology and statistics have a role to play, but they are not sufficient' to fully reveal the underlying dynamics (p. S122).

4. Quantitative measures: structural relationships as an IDR measure

Measurement of scientific output for the purposes of creating indicators is traditionally done using bibliometric approaches, so a large part of this review focused on bibliometrics. Bibliometrics have been refined over four decades based on the widely-held view that the scientific research process is incomplete without publication. Price (1978) argued that publication provides the function within science of correction, evaluation, and acceptance by a community. Published works are collected in journals, and a select set of journals is included in various databases such as the Web of Science (built and maintained by Thomson Scientific), or Scopus (built and maintained by Elsevier). These databases provide the raw materials used in current bibliometric efforts to measure IDR.

Bibliometrics are commonly used in science and technology policy analysis, evaluation, and the creation of indicators, as well as in the field of library and information science. The statistical procedures and analytical methods and tools have been refined along with the growth of scientific output. These methods were introduced by Derek de Solla Price and Eugene Garfield and adopted by the National Science Foundation in their Science & Engineering Indicators, Organization for Economic Cooperation and Development and other groups that track scientific output and productivity.[4]

Here, we consider different methods for measuring IDR suggested by the bibliometrics literature. A number of different methods are discussed; each one has benefits and drawbacks for fully representing IDR. There is no convergence in the literature on a single type of measure: multiple sources of data and measures are widely endorsed. In order to make any indicator useful for identifying and assessing the output of IDR, two questions must be addressed: (a) At what level of aggregation could and should measures be taken? and (b) At what level is knowledge input or output integrated? As noted at the outset of this review, the answers to these questions depend on the needs of indicator users, and differing needs will bring different datasets, analyses, and visualizations to bear.

Most IDR studies using bibliometrics draw upon Thomson Reuters products such as the Science and/or Social Science Citation Indexes and the Journal Citation Reports as their primary or only source of data. Several articles augment the ISI databases to draw from others such as Scopus (Adams, Jackson, & Marshall, 2007) or Medline (Boyack, 2004). Even in these cases, the additional databases have been mined to complement the ISI data. However, the current reliance on ISI can be seen as an historical path dependency (it was the only large-scale citation database for several decades). Scopus has now achieved high recognition and offers promise in that it currently indexes many thousands more source titles than the ISI databases. A variety of comparative studies of the two databases are now becoming available. In any rigorous analysis, having a standardized database is very important to the community's acceptance of the analysis.[5] We expect that any reliable indicator of IDR would use the most widely accepted database, and that indicators would prove to be reproducible and comparable over time. Some might consider Google Scholar (GS) to be an alternative data source for analysis, possibly simplifying or democratizing access to data. Nevertheless, at the time of this writing, GS is known to have deficiencies that make it less suitable for such studies than either Scopus or Web of Science (Bar-Ilan, Levene, & Lin, 2007). For example, it is known to have inconsistent coverage and duplicate records, it does not include metadata that would support disciplinary classification, and it does not support the type of bulk download that would be needed for large-scale studies.

The bibliometrics literature also reveals an underlying assumption about IDR. While explicit definitions differ, the authors see IDR as a phenomenon that emerges within (and because of) the dynamics of a larger knowledge system, which includes external drivers (such as the need to solve complex problems and funding priorities). While this is nearly universally acknowledged or implicitly assumed in the reviewed articles, two basic approaches emerge from the core literature: one that accounts for the larger system and one that does not. The articles that burrow down into the structure of science to examine relationships do so by placing relevant or desired data (authors, articles, or journals) into a hierarchical structure to allow comparisons.[6] Alternatively, some articles attempt to view the science system as a whole. These latter types use statistical relationships among key aspects of relevant or desired data to measure certain characteristics of aggregations of authors, articles, or journals: diversity, disparity, and balance.

Of the two approaches, the first approach, which we call a *structuralist* approach (subsuming both cognitive and social structures), is more frequently used. (The other approach which we call the *spatial* approach is described below.) Here we call

[4] Although the earlier seminal work is no longer cited, the core literature on measuring interdisciplinarity owes a great deal to the collected chapters in Elkana et al. (1978) – particularly the work of Price (1978), as well as to Garfield et al. (1978), in their chapters on metrics and on citation analysis, respectively.

[5] If a higher level, and yet very specific, framework could be developed onto which different databases could be mapped, then the choice of database would become less important.

[6] But note that *hierarchy* carries an inherent bias of prioritizing that is often challenged in forms of interdisciplinary work and the emergence of other publications. In addition, hierarchies are based on underlying assumptions about quality and importance.

Please cite this article in press as: Wagner, C. S., et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.06.004

attention, again, to alternative indicators that cut across while often challenging underlying assumptions about structure. In a dynamic knowledge system, *structure* is too static to capture all manifestations of intellectual work. The structure of science is variously characterized as consisting of: (a) the individual scientists or engineers; (b) groups of scientists organized around particular scientific topics, e.g. invisible colleges; (c) articles, notes, and letters published in refereed journals; (d) clusters of articles, e.g. co-citation clusters; (e) journals collecting the articles (i.e., *Journal of Theoretical Biology*); (f) specialties within sub-disciplines (e.g., applied bioinformatics); (g) sub-disciplines (i.e., bioinformatics); or (h) disciplines or fields of science (i.e., biology). At each stage beyond the individual scientist or engineer, a higher order of agglomeration is represented. In general, the reviewed articles choose among the levels of aggregation to conduct analysis; most often the level chosen is either articles or journals.

As Zitt (2005) points out, the literature does not converge on a universally accepted methodology or basis from which to uncover the structure of science (let alone how that structure may reveal IDR). Reliance on author affiliation, for example, is not widely used. A convenient but weak – and increasingly considered to be invalid – indicator of interdisciplinarity, as noted in Porter et al. (2007), is co-authorship of scholarly output. Among the many reasons why co-authorship is not used is that authors do not list their discipline/field in the published literature. While it is possible to examine *curricula vita* of authors to identify their discipline, this collection method requires expert judgment to place an individual into a specialty or discipline. It is also the case that researchers are often no longer working in the field in which they took their degree, making this process even weaker as a measure of IDR. The difficulty of using this measure has negated its application in most literature – only two of the articles we examined used individual authors as data: Porter, Roessner, and Heberger (2008) and Schummer (2004). Porter et al. drew upon a small sample of 110 researchers to compare alternative bibliometric indicators of the interdisciplinarity of their research outputs. Schummer notes that co-authorship analysis requires the tedious work of collecting data directly from papers, since the affiliation data do not exist separately.[7] This greatly reduces the usefulness of the option of using co-authors' affiliations as an IDR indicator.

The most common bibliometrics technique for measuring the output of IDR is citation analysis. A citation environment is defined as all journals that cite or are cited by a specific journal above a given threshold. Rafols and Meyer (2010) note that the percentage of citations outside of the discipline of the citing paper is the most common indicator of IDR, and indeed, this is the method suggested in the earliest literature by Garfield et al. (1978). Among the bibliometrics studies we reviewed, many use some form of citation or co-citation analysis (Adams et al., 2007; Boyack, 2004; Klein, 2008a; Levitt & Thelwall, 2008; Morillo, Bordons, & Gomez, 2001; Rinia, van Leeuwen, & van Raan, 2002; Zitt, Ramanana-Rahary, & Bassecoulard, 2005). Within citation analysis, the occurrence of what are considered discipline-specific citations pointing to other fields is assumed to reveal an exchange or integration among fields (an assumption that needs further examination, given the caveats associated with defining a discipline). Most of these articles seek to compare across disciplines at the level of institutions or disciplines.

Citation measures privilege *publication* as the major outcome of IDR. This is one of the sharpest limitations of this approach. Although the uses of citation-derived indicators are well accepted, a related question remains unexplored: In relying on the quantitative measures, what additional information about IDR impact is being lost? This may be exceedingly difficult to assess, as the extent of this gap is unknown at this time. Hamilton, Narin, and Olivastro (2005) and Zitt et al. (2005), among others, note that citation analysis is complicated by the fact that disciplines of science – however categorized – are vastly different in size. To address this problem, most producers of indicators have for decades offered relative measures (e.g., citation share over publication share) or ranking measures that facilitate comparisons between or among disciplines at a given level. The properties of various normalization techniques have to be explained when a measure at a given level (such as discipline) is intended to reflect structural discrepancies at lower levels (such as sub-field). Zitt offers a cross-scale perspective to address the aggregation problem in order to make indicators more rationalized across science.

Zitt et al. (2005) postulate that these broad families of citation measures are likely to provide different views of or different insights into the structure of science based upon the chosen level of analysis, which would be an expected effect given the differences in size of various fields. He presents a macro-approach to IDR at the article level (only articles were used, no notes or letters) to test the impact of the level of aggregation (measured by the number of citations to a unit of output) on the measure of interdisciplinarity (e.g., article, journal, discipline). Zitt draws upon the history of citation-based methodologies to present his refinement to the tool of co-citation analysis. His approach tests a number of levels of aggregation – journal, specialties (ISI subject categories), sub-disciplines, and disciplines – in order to test whether and how the level of aggregation affects the impact measure. (The disciplinary structure was based on the French Office of Science and Technology categories and the ISI subject categories.) Zitt modified the ISI subject categories in order to obtain strictly embedded levels without overlaps.[8] He then structured science into five levels: large-discipline (9 groups including multidisciplinary), sub-disciplines

---

[7] Scopus does code child affiliations (typically department level) where available in the original papers. A quick check of Scopus 2007 data shows that 69% of all author/affiliation records associated with individual articles contain child affiliation IDs. The number for the 2006 data is 68.5%. So, while not complete, Scopus does offer the possibility of using disciplinary affiliations to some degree.

[8] He notes that this strict categorization is not the best model of the organization of science, where disciplinary overlaps are a common feature, but rather a convenient simplification for the purposes of the study. 'Overlapping classes result in several normalized impact measures for each multi-assigned journal, which would obscure the proposed stylized analysis. The classification we use, and this is true for the comparable variants, does not claim to be

(31 groups), specialty (155 groups), and journals (3702 groups reduced to 3529 after filtering based on document type) for all articles drawn from the database for 1998.[9]

Van Raan (2005) notes that good citation analysis uses the same criterion to structure fields at various levels of aggregation. Underlying any judgment on the part of the analyst must be the understanding that citation practices are based on the choices made by scientists when they include citations in a paper – an important qualification, since it touches on the complex issue of what counts, or not. The action of centering the analysis on the judgment of the scientific practitioner makes this method more self-referential than other methods that might be used to indicate IDR (such as subject code analysis, which is several steps removed from the practitioner). The author's motives for giving a reference to a particular article may vary considerably, as Van Raan points out, but citation is the reference closest to the source of knowledge creation and thus to the integrative action.

As an example of the use of IDR measures at the journal level, Leydesdorff (2007b) describes the Journal Citation Reports (JCR) of the (Social) Science Citation Index, which contains structural information about citation-relation patterns of journals for each year. Matrices formed from these journal-journal citation data can be analyzed in terms of their structural dimensions using factor analysis. Additionally, graph-analytical approaches enable visualization of these data in terms of centrality measures. In the case of journal maps, the clusters can be designated in terms of scientific specialties. This enables analysis of IDR at the journal level by placing journals into disciplinary categories and then viewing the extent to which they have relationships with journals outside of that disciplinary category.

Leydesdorff notes that the Science Citation Indexes contain all the information needed to construct a huge matrix in which the cited journals provide information for the row vectors and citing journals for the column vectors (or vice versa). The resulting cells have values equal to the number of unique citation relations at the article level. The matrix is asymmetrical, and the main diagonal – representing within-journal citations – provides an outlier in the otherwise skewed distributions. As the similarity measure between the distributions for the various journals included in a citation environment, Leydesdorff uses the cosine between the two vectors, which is normalization to the geometrical mean. This technique can be used to measure IDR at the journal level as either an input (from the cited side, measuring its contributions to or impact on other journals/disciplines) or output (from the citing side, measuring its uptake or use of information from other journals/disciplines).[10] When the threshold is set equal to zero, it is possible to map the complete citation context of a journal, showing its citation environment within a discipline and outside a discipline (based upon certain assumptions about disciplinary orientation of journals). This set of relationships can be represented in tables or in abstract space based on vector analysis, showing similarities (or linkages) and differences (or variations). This latter process brings our discussion to the point of considering using a metaphorical knowledge space to conduct IDR analysis.

5. Spatial distances as an assessment tool

A second approach to using bibliometrics is a methodology that describes a landscape, or space within which science operates, typically from the point of view of a single object (journal, paper, or author). As noted above, spatial distance as a measure to analyze and visualize IDR was suggested by Garfield et al. (1978). The usefulness of this measure has been enhanced by recent developments in computing and algorithms that can be used to standardize the analysis and to bring into view the underlying dynamics of the relationships among and across disciplines. Many recent studies have employed spatial distance approaches alone or in combination with other tools to understand and measure IDR (Boyack, 2004; Hamilton et al., 2005; Leydesdorff & Schank, 2008; Leydesdorff, 2007a, 2007b; Porter & Rafols, 2009; Rafols & Meyer, 2010; Stirling, 2007; Van den Besselaar & Heimeriks, 2001; Van Raan, 2005). The articles referenced here are not a comprehensive list, but are representative of our view of the best of current approaches. Spatial approaches are often complemented by advances in visualization and animation techniques that have placed the problem of distinguishing structural change based on differences (variation) once again on the agenda. One classical method is to use factor analysis in which each factor can be labeled as a category or *discipline*. Then, the elements of the matrix that load upon multiple factors above some preset threshold are considered the most interdisciplinary (Van den Besselaar & Leydesdorff, 1996; Van den Besselaar & Heimeriks, 2001). Factor analysis defines an emergent structure from data based on a set of distances, rather than using a pre-imposed structure. Beyond factor analysis, the following sections explore three types of spatial approaches: (a) classification and taxonomy of science to identify diverse aspects of the system; (b) a description and measurement of entropy within the system as a measure of diversity; and (c) a measure of the centrality of specific parts of the knowledge network as an indicator of IDR activities.

5.1. Diversity

Diversity measures are suggested by Rafols and Meyer (2010), Stirling (2007), and Porter and Rafols (2009). Several papers propose adapting diversity and coherence concepts to IDR. (We acknowledge the expanded range of indicators in

an accurate representation of the structure of science. It is, rather, a limited but realistic tool suitable for macro-analysis purposes. (p. 376)'.

[9] Disciplines at different degrees of size and formality also need to be factored in as well as emergent fields and, in the latter case, of sufficient recognition that they are entering into revised taxonomies used by the National Research Council, National Institutes of Health, and other funding and policy agents.

[10] We note that common usage of the terms multi-, inter-, and trans-disciplinary rarely distinguish between the input and output directions of IDR as described here.

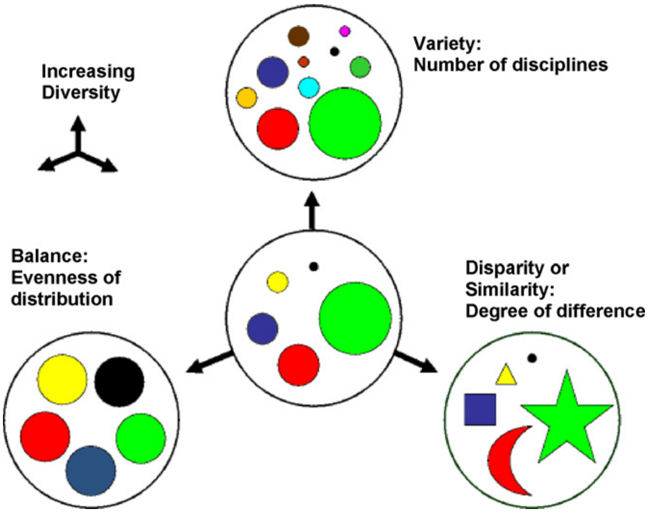

Fig. 2. Schematic representation of the attributes of diversity, from based on Stirling's conceptualization (2007).

the discourse on evaluating IDR, which by default incorporates the topic of measurement and metrics. It also engages the need for heuristics and, in the case of the Stokols et al. (2008) model, feedback to multiple and diverse outcomes.) Rafols and Meyer introduce disciplinary diversity indicators to describe the heterogeneity of a bibliometric set viewed from pre-defined categories; that is, using a top-down (or structural) approach that locates each of the elements (e.g. papers) of the set on one of the categories of the global map of science. A schematic of their definition of diversity is shown in Fig. 2. In their approach, a network is constructed out of the relationships among the elements. From this data set, a network coherence measure is derived. The measure depicts the intensity of similarity relations within a bibliometrics set using a bottom-up approach (e.g., co-citation). This allows the network relationship to reveal the structural consistency (or lack thereof) of the publication network. Rafols and Meyer carry out case studies on individual articles in bionanoscience to illustrate how these two perspectives identify different aspects of interdisciplinarity: disciplinary diversity indicates the large-scale breadth of the knowledge base of a publication (the multidisciplinary scope); changes from lower to higher network coherence reflect knowledge integration. They suggest that the combination of these two approaches may be useful for comparative studies of emergent scientific and technological fields where new and controversial categorizations are accompanied by equally contested claims of novelty and interdisciplinarity (Chen et al., 2009). Theoretical interdisciplinarity and methodological interdisciplinarity need to be carefully disentangled to pursue this further.

Several other papers suggest combining similarity and distance measures to derive an indicator. Porter et al. (2007) developed measures of integration and specialization of research outputs to understand where IDR is within the system of science. The integration indicator measures cognitive distance (dispersion) among the subject categories of journals cited in a body of research. The specialization indicator measures the spread of the journal subject categories in which a body of research is published. Measures using similarity and variation are also used to compute network properties such as centralities, a measure proposed by Leydesdorff (2007a).

5.2. Entropy

Entropy is a mathematical concept drawn variously from thermodynamics, statistical mechanics, and information theory. It is a measure of disorder or uncertainty in a system – in this case the system of science. Entropy is a particular case of the general concept of diversity (Stirling, 2007) and hence similar to the index described in Section 5.1, and can be used to measure either inputs or outputs. Hamilton et al. (2005), use it as an indicator of disciplinary impact by focusing on the intensity of knowledge streams between research fields, a feature it shares with Van den Besselaar and Heimeriks (2001). This is done from an input, or forward-looking perspective (in contrast to the majority of studies referenced in this review, which are output, or backward-looking in nature) in which the system is the network of citations coming from all other fields to one field. Since Hamilton et al. (2005) are primarily interested in the impact one field has on all other fields, they ignore citations coming from within the field. For simplicity's sake, they divide the system into nine total fields. The premise is described in the following way:

> If we hold up a paper and say 'This paper has cited Mathematics', then how much information do we need to determine the field of the citing paper–i.e., how many yes/no questions do we need to ask? The answer to that question is the amount of uncertainty, or entropy, in the system. (p. 5)

Table 2
) table of units of analysis as related to network analysis and.

Unit of analysis	Questions related to			Commonly used algorithms
	Fields and paradigms	Communities and networks	Research performance or competitive advantage	
Authors		Social structure, intellectual structure, some dynamics	Use network characteristics as indicators	Social network packages, multidimensional scaling, factor analysis, Pathfinder networks
Documents	Field structure, dynamics, paradigm development		Use field mapping with Indicators	Co-citation, co-term, vector space, latent semantic analysis, principle components analysis, various clustering methods
Journals	Science structure, dynamics, classification, diffusion between fields			Co-citation, Intercitation
Words		Cognitive structure, dynamics		Vector space, latent semantic analysis, latent dirichlet allocation
Indicators and metrics			Comparisons of fields Institutions, countries, etc., input–output	Counts, correlations

The answer they propose is a distribution of citations across the external fields of the system. At one extreme, maximum multidisciplinarity (in their definition of the term), there would be a situation of equal distribution in which each field has the same number of citations going to one field.

5.3. Betweenness centrality and boundary spanning

Leydesdorff (2007a) has suggested that betweenness centrality can be used as a measure of interdisciplinarity at the journal level. Schummer (2004) suggests using a similar measure at the individual author level. This variation measure uses an integration of multivariate and time-series analysis to view the relationships among different parts of a knowledge system. By measuring the vector space between nodes in a citation environment, and with the reference point being the system as a whole, the measure of those journals that are most *between* others in the system can be identified, measured, and visualized. In related work, Leydesdorff (2007b) also explores the possibility of examining the position of the nodes across the network. The nodes that link other nodes are less connected to each other but are also the points doing the most boundary spanning.

In order to place the node within the knowledge system, analysis requires a map of all the knowledge space in which a journal is located. This measurement of the system as a whole has been the goal of numerous recent studies (e.g., Boyack, 2004; Leydesdorff & Rafols, 2009; Morillo, Bordons, & Gomez, 2003). Boyack notes that most mapping applications are used for research assessment or to structure the dynamics of disciplines. Table 2 shows the different types of mapping tools that Boyack reviews and how they can be used, what they reveal in terms of dynamics, and the commonly used algorithms applied to different levels of aggregation.[11]

Using factor analysis, Leydesdorff and Rafols (2009) analyzed the full set of ISI subject categories to find that it can be decomposed into 14 factors or macro-disciplines. Each of these macro-disciplines can be clustered, with the clusters providing spaces of high population amidst open spaces. Links across these spaces show the interdisciplinary character of the populated spaces. This system-wide mapping of population spaces and their links provides a basis for determining the extent to which different fields are more or less interdisciplinary in character.[12] Morillo et al. (2003) attempt to establish a tentative typology of disciplines according to their degree of interdisciplinarity measured through multi-assignation of journals to ISI subject categories, by which 'the assignment of a journal to more than one category indicates the existence of cognitive links between disciplines, which can result in interdisciplinary research' (p. 1238).

5.4. Indices

Due to the limitations of any single measure of IDR, it is worth exploring the development of one or more indexes of interdisciplinarity. (We anticipate that any index of interdisciplinarity would extend beyond citation impact.) Adams et al.

[11] Boyack's classifications depend on underlying assumptions about indicators used to map fields. The NRC indicated a number of years ago that in some areas of science the indicators are shifts in language, an observation with powerful implications for thinking about keywords and other searching strategies.
[12] The problem of this approach is that it is dependent on the particular classification used. This is because disciplinary classifications, such as ISI SCs, differ as much as 50% in the attribution of journals into categories. However, thanks to statistical aggregation, surprisingly similar maps of science can be derived using different classifications (Rafols & Leydesdorff, 2009). Indeed, Klavans and Boyack (2009) have shown that most of global mapping studies agree on the core structure of science. This suggests that although the science mapping approach may be very dependent on the classification used at a fine-level, at a coarse level these differences not to be relevant.

Please cite this article in press as: Wagner, C. S., et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.06.004

(2007) suggest that the index could be done based on Thomson's multidisciplinary category, which is drawn from the journal classification scheme. The Adams et al. article was written for a specific policy application (i.e., the UK government's impact analysis at the university level). Though the index idea is a good one, the Adams et al. index is too narrow to serve as a broader indicator of IDR at the system-wide level.

As discussed in Section 5.1, Porter and Rafols (2009) also suggest a diversity index as a way to incorporate multiple measures of IDR. Their index measures how integrative of fields particular research articles are based on the association of the journals they cite to corresponding subject categories of the Web of Science. This is a valuable perspective but indicators of integration are far more plural. They present a practical way to map scientific outputs, again based on dispersion across subject categories. Following Stirling's heuristic, they argue that in order to reveal interdisciplinarity, multiple dimensions must be included. They suggest that these are (a) the number of disciplines cited, or *variety*, (b) the distribution of citations among disciplines, or *balance*, and (c) how similar or dissimilar these categories are, or *disparity*. This is illustrated in Fig. 2. The key feature of this integration score (also called the Rao–Stirling diversity index) is that it captures not only the number of disciplines cited by a paper and their degree of concentration (as Herfindahl or Shannon indices do), but also provides a measure of how disparate (i.e., how different) are these disciplines. In order to do so, it relies on a specific metric of distances (or similarities) between pairs of disciplines.

6. Limitations and additional issues

The use of bibliometrics tools for the measurement of IDR skews measures towards use of indexed literature in bibliographic databases. On one level, this appears to be a limitation given that much of IDR takes place as a dynamic process operating at a number of levels (whether one considers the social, the cognitive, or the knowledge dynamics of the measurement process). Output measures do not account for these dynamic processes, thus they are narrow to the point of offering what some consider a distorted view of IDR within the sciences. Nevertheless, knowledge creation has always been more dynamic than can be represented quantitatively – progress in measurement has pushed the possibilities for measurement to incorporate new approaches, which are discussed in the next section.

6.1. Classification-based vs. bottom-up approaches

Most current bibliometric-based measures of IDR output rely largely on journal categories established by ISI. However, we note limitations in the use of journal categories for measuring IDR because of the dependence upon a pre-defined taxonomy or category structure. Studies using journal categories are often viewed as biased due to a lack of consensus around the accuracy of any particular journal category system. Classification-based measures can be useful as a first start, especially when used to compare large areas of science using large amounts of data. Nevertheless, issues around the use of underlying taxonomies or classification schemes as the basis for IDR measures is highly problematic. With no consensus on the best categorization and considerable evidence that various measures of IDR yield quite different results depending on the classification system chosen for analysis, it is apparent that problems remain for any IDR measure based on a pre-defined classification scheme.

Bottom-up approaches based on the clusters formed by articles, using co-citation (Marshakova, 1973; Small, 1973), co-words (Callon & Law, 1983), or bibliographic coupling (Kessler, 1963), can capture knowledge integration *in the making*, evidenced by people working on new problems that span previously formed clusters (i.e., crystallization, Leydesdorff's (2007a) or Chen's (2004) betweenness centrality (Freeman, 1977), Rafols and Meyer's (2010) network coherence or value degree centrality). Each approach may tell a useful story. This structural approach is better at capturing emerging developments that do not fit into existing categories, while the classification-based approach might be useful at large-scale explorations, such as the disciplinary breadth of universities.

6.2. Static vs. dynamic measures of IDR

There are serious implications for measurement of IDR posed by the fact that IDR itself is a dynamic, moving target. What is currently thought of as a highly interdisciplinary field is a point-in-time perception of how far apart the present constituent categories were at an earlier time, suggesting that a measure of the distance between two topics or between fields of science should be based on the analysis of large amounts of data over time. Moreover, interdisciplinary practices can be assumed to differ across disciplines. For these reasons, IDR could be considered as a process, an evolution, rather than a state, thereby requiring that – if its characteristics are to be captured – they should portray development over time. In fact, given the consensus that a central concept of IDR is a process of knowledge integration, this in itself suggests a dynamic process leading from less to more integration. While an existing field might be characterized by a static IDR measure as 'highly integrated', in the process of evolving to its highly integrated state it may have lost much of the novelty and breakthrough value often associated with incipient IDR. Loet Leydesdorff, in an e-mail to us,[13] asked whether articles in hindsight might

[13] Personal communication, 16 March 2009.

be considered very interdisciplinary, though they might not have been so perceived upon immediate publication; indeed, the converse might also be the case. In any event, current IDR measurement techniques are static in nature and use a single finite time window, like looking at a one-dimensional representation of a physical structure.

6.3. Limitations of existing databases as sources of IDR measures

Existing databases have a number of limitations that raise questions about their use for developing bibliometric-based measures of interdisciplinary research. Several related issues are important for future assessments of IDR. The NSF (and other policy-oriented bodies interested in assessing IDR) seeks to support research that addresses complex scientific and social problems. Doing so requires the integrated contributions of varied fields of knowledge. Indeed, one of the most important rationales for the support and measurement of IDR is to help tackle problems where social and scientific issues are intermeshed. So in principle, assessments and conduct of IDR must include the humanities and social sciences as well as the physical and medical sciences. However, developing meaningful measures of IDR based on assumptions that apply only to the physical and medical sciences (such as the significance of the order of authorship of journal articles which itself varies within the physical and medical sciences), the use of standardized databases such as the Journal Citation Reports limit development of highly valid measures of IDR. Reliable data to assess developments in the social sciences and humanities do not exist because current bibliometric databases (ISI, Scopus) do not cover books, book chapters and many regional non-English journals in which authors from the social sciences, humanities, and some applied fields mainly publish. Data in the humanities are harder to obtain, and links between the humanities and social/physical/medical sciences are much less explicitly codified than those within physical and medical sciences. Thus, given the limits of bibliometrics, we may be missing the most socially relevant IDR interactions because of limitations in the current contents of bibliographic databases.

7. Observations

The literature reviewed here takes many approaches to understanding and measuring interdisciplinary research. Several points of consensus emerge. First is the focus on integration of knowledge as an essential factor distinguishing interdisciplinary research. Most authors presume that science is structured (explicitly or implicitly) around a problem/facts/methodology nexus, often labeled as *research specialty, discipline* or *knowledge domain*. The literature reviewed here, complemented by the National Academies panel report (2005), suggests a definition focused on integration of expertise, theory, method, or data from at least two different disciplines of research, resulting in new collaborations and published outputs that reflect new insights into, or understanding of, a problem, a method, a data set, or phenomenon.

Among those relying on bibliometrics techniques, it is clear that citation analysis in several forms is widely accepted as a basis upon which to develop measures of IDR in various groupings of research output, but it is not the only measure. Within this set of tools, there are many methods used to reveal the structure or variety of knowledge sources within science. Exploration of various measures will continue to be of interest to academic bibliometrics researchers. But for the purposes of indicator development for use in policymaking and research management, such a variety of approaches to IDR become a complication rather than a revelation. Assessment of interdisciplinary research inputs, processes, outputs, and outcomes is still a work in process. Similarly, it is clear that there are very few studies that link input and output measures – a measurement challenge that remains to be probed in the future. Although previous mentions of inputs referred to citing direction, attitudes, or allocation or resources, it may be important to expand the definition of inputs to include funding, regulation, and policy.

Fundamental questions surrounding IDR need to be addressed before agreement can be reached on the most appropriate measures addressing different user needs. Among them:

1. How do different levels of aggregation of research activity (people, teams, institutions, countries or geolocation) affect selected measures of interdisciplinary research?
2. What is the appropriate unit of analysis for addressing different questions about interdisciplinary research (e.g., papers, journals, fields of science, authors, departments, projects)?
3. What types of measures are best for computing the inputs, processes, and outputs of interdisciplinary research (e.g., simple counts, network measures, dynamic models, or a combination)?
4. What types of normalization (necessary to account for the different relative sizes and dynamics of science across research fields) are required to match the different units of analysis and granularity listed above?
5. What are the contexts and processes that foster knowledge integration in research, and how best can the level of integration be identified and measured, and linked to desired outcomes, however defined?
6. What combination of qualitative and quantitative measures is most revealing of the actual working processes and the outputs of IDR?

In addition, work to link broader processes and outcomes, such as those measured by Stokols and others, with features of bibliometrics data, would also be very useful. If those broader processes could be correlated with bibliometrics data, outcomes as well as outputs could be scaled. At the same time, potential users of IDR measures could work with the research community, articulating their needs in the form of questions whose answers require reliable and valid measures of interdisciplinary research.

Among the various approaches examined, little common ground can be found between those who study IDR within and resulting from social or teaming processes and those who understand the knowledge system as independent of the social system. Those studying IDR as a teaming process seek measures that reflect the underlying social dynamics and/or motives of the researchers. This group is farther from agreeing on indicators than are those who view the knowledge system as having an internal dynamic separate from the social processes or from the intentions of the team members. Those researchers are more likely to view as valid the quantitative indicators, particularly the diversity and betweenness measures, as representative of IDR. The literature does not offer a bridge between these two approaches, and it may be that the different camps will never agree on appropriate metrics – barring a common agreement on a theory of knowledge creation.

As a final word we suggest, at the least, a need for the emergence of a process using a combination of articles and citations in a bottom-up approach. Among the articles reviewed here, several suggest measures that can be used to create indicators in the near term, and the level of creativity being applied to analysis of communications is encouraging. This is particularly true of the articles that offer multiple measures combined in some way, such as those being placed into indices or measured and then mapped so as to incorporate both similarity and variation in the analysis. Which of these measures are most appropriate for addressing particular questions concerning the interdisciplinary content of research output, at what levels of aggregation, and with what degree of validity and reliability, must be explored in further research and discussion.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Foundation for this review, particularly the guidance of Lawrence Burton of the Social Behavioral, and Economics Directorate. We also thank several reviewers and commentators, especially Loet Leydesdorff, Alan Porter, and Dan Stokols. The work by Katy Börner is partially supported by the James S. McDonnell Foundation.

References

Adams, J., Jackson, L., & Marshall, S. (2007). *Report to the Higher Education Funding Council for England: Bibliometric analysis of interdisciplinary research*. Leeds, UK: Evidence, Ltd.

Bar-Ilan, J., Levene, M., & Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics*, *1*, 26–34.

Boulding, K. E. (1956). General systems theory: The skeleton of science. *Management Science*, *2*, 197–208.

Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5192–5199.

Callon, M., & Law, J. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*, 191–235.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5303–5310.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, *3*(3), 191–209.

Croyle, R. T. (2008). The National Cancer Institute's transdisciplinary centers initiatives and the need for building a science of team science. *American Journal of Preventive Medicine*, *35*, S90–S93.

Darden, L., & Maull, N. (March, 1977). Interfield Theories. *Philosophy of Science*, *44*(1), 43–64.

Elkana, Y., Lederberg, J., Merton, R., Thackray, A., & Zuckerman, H. (Eds.). (1978). *Toward a metric of science: The advent of science indicators*. New York: John Wiley.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*, 35–41.

Garfield, E., Malin, M., & Small, H. (1978). Citation data as science indicators. In Y. Elkana, J. Lederberg, R. Merton, A. Thackray, & H. Zuckerman (Eds.), *Toward a metric of science: The advent of science indicators* (pp. 179–208). New York: John Wiley.

Gray, B. (2008). Enhancing transdisciplinary research through collaborative leadership. *American Journal of Preventive Medicine*, *35*, S124–S132.

Hall, K. L., Feng, A. X., Moser, R. P., Stokols, D., & Taylor, B. K. (2008). Moving the science of team science forward. *American Journal of Preventive Medicine*, *35*, S243–S249.

Hall, K. L., Stokols, D., Moser, R. P., Taylor, B. K., Thornquist, M. D., Nebeling, L. C., et al. (2008). The collaboration readiness of transdisciplinary research teams and centers. *American Journal of Preventive Medicine*, *35*, S161–S172.

Hamilton, K. S., Narin, F., & Olivastro, D. (2005). *Using bibliometrics to measure multidisciplinarity*. Westmont, NJ: ipIQ, Inc.

Kessel, F., & Rosenfeld, P. L. (2008). Toward transdisciplinary research: Historical and contemporary perspectives. *American Journal of Preventive Medicine*, *35*, S225–S234.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*, 10–25.

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, *60*, 455–476.

Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities*. Charlottesville, VA: University Press of Virginia.

Klein, J. T. (2008a). Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine*, *35*, S116–S123.

Klein, J. T. (2008b). The rhetoric of interdisciplinarity. In A. Lunsford, K. Wilson, & R. Ebeerly (Eds.), *The Sage handbook of rhetorical studies* (pp. 265–284). Thousand Oaks, CA: Sage.

Levitt, J. M., & Thelwall, M. (2008). Is multidisciplinary research more highly cited? A macrolevel study. *Journal of the American Society for Information Science and Technology*, *59*, 1973–1984.

Leydesdorff, L. (2007a). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*, 1303–1319.

Leydesdorff, L. (2007b). Mapping interdisciplinarity at the interfaces between the Science Citation Index and the Social Science Citation Index. *Scientometrics*, *71*, 391–405.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, *60*, 348–362.

Leydesdorff, L., & Schank, T. (2008). Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments. *Journal of the American Society for Information Science and Technology.*, *59*, 1810–1818.

Marshakova, I. V. (1973). A system of document connections based on references. *Scientific and Technical Information Serial of VINITI*, *6*, 3–8.

Mâsse, L. C., Moser, R. P., Stokols, D., Taylor, B. K., Marcus, S. E., Morgan, G. D., et al. (2008). Measuring collaboration and transdisciplinary integration in team science. *American Journal of Preventive Medicine*, *35*, S151–S160.

Miller, M., & Mansilla, V. B. (2004). *Thinking across perspectives and disciplines*. Boston, MA: Harvard Graduate School of Education.

Mitrany, M., & Stokols, D. (2005). Gauging the transdisciplinary qualities and outcomes of doctoral training programs. *Journal of Planning Education and Research*, *24*, 437–449.

Morillo, F., Bordons, M., & Gomez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, *51*, 203–222.

Morillo, F., Bordons, M., & Gomez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, *54*, 1237–1249.

Nash, J. M. (2008). Transdisciplinary training: Key components and prerequisites for success. *American Journal of Preventive Medicine*, *35*, S133–S140.

Porter, A. L., Cohen, A. S., Roessner, D. J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, *72*, 117–147.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*, 719–745.

Porter, A. L., Roessner, D. J., Cohen, A. S., & Perreault, M. (2006). Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation*, *15*, 187–195.

Porter, A. L., Roessner, D. J., & Heberger, A. E. (2008). How interdisciplinary is a given body of research? *Research Evaluation*, *17*, 273–282.

Price, D. J. D. (1963). *Little science, big science*. New York: Columbia University Press.

Price, D. J. D. (1978). Toward a model for science indicators. In Y. Elkana, J. Lederberg, R. Merton, A. Thackray, & H. Zuckerman (Eds.), *Toward a metric of science: The advent of science indicators.* (pp. 69–95). New York: John Wiley.

Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. *Journal of the American Society for Information Science and Technology*, *60*, 1823–1835.

Rafols, I., & Meyer, M. (2010). Diversity measures and network centralities as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*, 263–287.

Rhoten, D., Caruso, D., & Parker, A. (2003). *A multi-method analysis of the social and technical conditions for interdisciplinary collaboration*. Hybrid Vigor Institute: NSF BCS-0129573 Final Report.

Rinia, E. J., van Leeuwen, T. N., & van Raan, A. F. J. (2002). Impact measures of interdisciplinary research in physics. *Scientometrics*, *53*, 241–248.

Rosenfield, P. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science of Medicine*, *35*, 1343–1357.

Schmidt, J. C. (2008). Towards a philosophy of interdisciplinarity: An attempt to provide a classification and clarification. *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, *5*, 53–69.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, *59*, 425–465.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*, 265–269.

Stirling, A. (2007). A general framework for analyzing diversity in science, technology and society. *Journal of the Royal Society*, *4*, 707–719.

Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., et al. (2003a). Evaluating transdisciplinary science. *Nicotine & Tobacco Research*, *5*, S21–S39.

Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., et al. (2003b). Evaluating transdisciplinary science. *Nicotine & Tobacco Research*, *5*(Suppl. 1), S21–S39.

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*, *35*, S77–S89.

Stokols, D., Misra, S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008). The ecology of team science: Understanding contextual influences on transdisciplinary collaboration. *American Journal of Preventive Medicine*, *35*, S96-S115.

The National Academies. (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press.

Van den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. *Journal of the American Society for Information Science*, *47*, 415–436.

Van den Besselaar, P., & Heimeriks, G. (2001). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In *Proceedings of the 8th international conference on scientometrics and informetrics* (pp. 705–716).

Van Raan, A. F. J. (2005). Measurement of central aspects of scientific research: Performance, interdisciplinarity, structure. *Measurement*, *3*, 1–19.

Wagner, C. S., Roessner, J. D., & Bobb, K. (May, 2009). *Evaluating the output of interdisciplinary scientific research: A review of the literature*. Arlington, VA: SRI International.

Zitt, M. (2005). Facing diversity of science: A challenge for bibliometric indicators. *Measurement*, *3*, 38–49.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, *63*, 373–401.

Please cite this article in press as: Wagner, C. S., et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics* (2010), doi:10.1016/j.joi.2010.06.004