Research Evaluation, 19(1), March 2010, pages 19–27
DOI: 10.3152/095820210X492503; http://www.ingentaconnect.com/content/beech/rev

Identification of converging research areas using publication and citation data

Reindert K Buter, Ed C M Noyons and Anthony F J van Raan

Converging research is the emergence of new interdisciplinary research from fields which showed limited mutual interdisciplinary connections before. We describe three search strategies to identify converging research using data extracted from the WoS, including the social sciences and humanities. The field-to-field references (FFR) strategy uses citations from one journal subject category (JSC), to another; the keyword sets (KWS) strategy tracks the co-occurrence of keywords from different JSCs; and the affiliation patterns (AFP) strategy traces the co-occurrence of fields of research in author affiliations of papers. Resulting publication sets were assessed using data such as journal names, titles of publications, and titles of cited publications. Experts validated nine converging research areas that were detected using the KWS and FFR strategies; none were found with AFP strategy.

IN FEBRUARY 2008, the *Lancet* published an article on the outcomes of a trial which studied the effects of a treatment with probiotics on patients suffering from severe acute pancreatitis (Besselink *et al*, 2008). The most significant outcome of this study was an increased risk for mortality in the group that was supplied the probiotics, and the study was stopped prematurely. In the same issue, the *Lancet* published an editorial bearing the suggestive title ‘Probiotics or con’, which was devoted to this grave outcome and its larger consequences. The editorial concluded that without further research, it was no longer tenable to regard the use of probiotics in consumer products as completely risk-free.

This trial and its editorial response illustrate an interesting interdisciplinary phenomenon. First, probiotics, which are well-known food additives, were included in the treatment of severely ill people. Next, following the suggestion of the *Lancet* editorial, probiotics as food additives should now be scrutinized due to the outcome of this application. In a sense, this would complete an interdisciplinary circle, where a result (probiotics) from one field (food science and technology) influences research (trial study) in another field (clinical medicine) and vice versa.

Developments such as this, where distinct research areas start to apply problems and tools from one another, possibly leading to new research directions or even new research areas, are considered examples of *converging research* in this article. More formally, we define converging research as the emergence of new interdisciplinary research from fields which showed limited mutual interdisciplinary connections before. In this definition, emergence is understood as the appearance of thematically related research which shows a significant growth in the number of publications. Also, the converging research is the result of research in two areas which had limited interdisciplinary connections before, and is thus both new interdisciplinary research, and a novel combination of disciplines. Our definition explicitly excludes intra-disciplinary emergence (i.e. within fields), and emergence of this type was not considered in our research.

The publication of the report on the NSF/DOC-sponsored workshop ‘Converging Technologies for Improving Human Performance’ (Roco and Bainbridge, 2003), was important for the widespread introduction of the notion of converging research.

Reindert K Buter (corresponding author), Ed C M Noyons and Anthony F J van Raan are at the Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62a, P.O. Box 905, 2300 AX Leiden, The Netherlands; Email: buter@cwts.leidenuniv.nl; Tel: +31 (0)71 527 3904.

There, it was presented as a conceptual framework based on a 'unification of science in nature' which allows the creation of mutually enabling research areas with 'synergistic effects' (Roco and Bainbridge, 2003: 1). The workshop focused on the four fields: *Nanotechnology*, *Biotechnology*, *Information sciences* and *Cognitive sciences* (the NBIC fields), and the application to the improvement of human performance in general.

The European response to this workshop was the assembly of an expert group. This group published a report (Nordmann, 2004) titled 'Converging Technologies for a Knowledge Society' (CTEKS). This report stated that converging research was not limited to the NBIC fields, but could originate from anywhere in science, and that the application was generalised to current societal, economical and scientific problems (instead of the improvement of human performance). However, *both* the NBIC and the CTEKS reports stressed the importance of the research in the social sciences and humanities to investigate aspects such as social and ethical consequences of the application of the research.

The possibility of applying research developments on current problems shows the potential use converging research can have for policy-makers. However, just what sets converging research apart from other interdisciplinary developments is not always clear. Expanding our knowledge about this issue was one of the objectives of an EU project titled 'New Emerging/Converging Clusters of Science and Technology'. The project's most important goal was to identify and describe converging research addressing current societal, economical or scientific problems using publication data. Even more, these data should not only focus on well-covered fields such as the life sciences or the material sciences, but also include those from social sciences and humanities. In this article we will review the methodology developed during the EMCOTEC project and reflect upon the results as well as on some of the design choices made.

Converging research is defined as a case of emerging research of a novel, interdisciplinary nature. Many examples exist of descriptions and analyses of emerging research areas with both disciplinary and interdisciplinary roots, for example mathematical logic (Berg and Wagner-Döbler, 1996), superstring theory (Budd and Hurt, 1991), elements of spectroscopy (Ródenas-Torralba *et al*, 2006), bioelectronics (Hinze, 1994) or nanoscience and nanotechnology (Schummer, 2004). Also, general methodologies to find emerging patterns exists, such as those described by Morris (2005), Shibata *et al* (2008) or by Takeda and Kajikawa (2009). However, like the descriptions of the emerging research areas, these general approaches are applied to a limited set of publications or with a specific topic in mind. Moreover, none of these publications deals with the general and broad search for converging research as defined above.

Data and method

This article presents an approach to find converging research. The development of this approach faced some important challenges:

- The many possibilities to combine research from different parts of science;
- The many ways in which one area can influence another, such as cognitively, socially or methodologically;
- The subjective aspects of core notions such as 'research area', or 'interdisciplinary connections';
- The difficulty in generalising known examples of converging research, so that finding one does not necessarily help finding others;
- The unknown number of examples of converging research, making it very difficult to quantify search performance in terms such as recall and precision.

The resulting approach is a combination of both quantitative and qualitative aspects, which we will explain in more detail below.

The quantitative aspects use publication and citation data extracted from the Web of Science (WoS) database, which is available to the CWTS under license from Thomson Reuters. The publication data consisted of elements such as source (journal) names, titles and abstracts, affiliations, and keywords for publication published between 1990 and 2005. The citation data consisted of all citations from these publications to publications also covered by the WoS.[1]

We identified converging research in a three-step process: search, inspection and validation by experts. Search was performed by using three strategies described below, which differed in the representation of research field and interdisciplinary connections: the field-to-field references approach; the keywords-sets approach; and the affiliation patterns approach. All strategies tested the development of interdisciplinary connections between pairs of fields. The selected pairs were inspected by using additional publication data to understand what the potentially converging research between these pairs was about. Finally, pairs which appeared to represent converging research, were subsequently shown to experts and only accepted if they also recognised it as such. Below we explain the individual approaches in more detail, together with basic notions about growth and size. As will become clear, the strategies show variation in the implementation of these basic notions, because details of the approaches were developed and fine-tuned during the project, and experience gained from the application of one strategy was used in the fine-tuning of the next.

The above definition makes convergence a case of (interdisciplinary) emergence. Therefore, a notion of significant growth is needed. In our project, we focused on fast, non-linear growth, since this type has in many studies been associated with emerging research themes (Tabah, 1999; Goffman, 1971 and May, 1966). One of the tools we developed to identify such growth, which we called the relative total growth (RTG), is the absolute difference between the number of papers in the first and last year, divided by the total number of papers. Since smaller values of the RTG are due either to a small difference between the first and last years, or to a small change relative to the total number of papers, important examples of fast growth can be found at the top of a list sorted on descending RTG. However, this indicator is not perfect, and for example does not take into account the shape of the development between the first and the last year; therefore, the RTG was used only to sort promising results.

A research area should show a basic, noticeable level of activity, which should also be measurable in number of publications. However, it is very difficult to quantify ‘a basic level’ exactly. Looking at examples of emerging research areas described in the literature, a total number of publications published per year at the end of the development, can be for example around 100 for fuzzy set theory (Berg and Wagner-Döbler, 1996), or around 250 for bioelectronics (Hinze, 1994). Moreover, the number of publications involved in earlier years can be small: even fewer than 10 publications per year (Hinze, 1994). The search strategies described below used different lower bounds on the number of publications, due to the experience gained in applying earlier strategies.

We now describe the three strategies in more detail. The most important difference between these strategies is the way they characterise research fields (sets of papers with related research), as well as interdisciplinary connections (use of research from different areas). In the different characterisations, we have tried to span different dimensions of the science system: references represent (among others) the social and cognitive dimensions, the social and geographical dimensions, and keywords the cognitive and vocabulary dimensions.

The most important difference between the three strategies we describe is the way they characterise research fields (sets of papers with related research), as well as interdisciplinary connections (use of research from different areas)

Field-to-field references

The first strategy we developed is the field-to-field references strategy (FFR). It characterises research fields as journal subject categories (JSCs), which assign each journal in the WoS to at least one and at most six categories. These JSCs are part of the WoS and at the time our research was conducted, there were 243 JSCs. Although this classification is not perfect, it provides a clear, fixed and consistent field definition suitable for automated procedures (Van Raan, 2008). Nevertheless, problems exist, especially with the differences in size and (breadth of) scope, and the multiple assignments of journals to different JSCs. However, such multiple assignments can also be regarded as hubs between categories, which was used by Morillo *et al* (2003) to develop an indicator for interdisciplinarity.

The FFR strategy characterised interdisciplinary connections by the references from one category to another. To give an interesting example of this, the JSCs *Religion* and *Allergy* are connected through a citation in 2000 from the journal *Zeitschrift für Evangelische Ethik* to the journal *International Archives of Allergy and Immunology* by Haniel (2000) citing Hammer *et al* (1998). Our strategy excludes author self-citations and journal self-citations, since these may be indicative for other processes than use or exchange of knowledge (De Solla Price, 1981). Our analysis covered 15 years of publication data (1991–2005). The lower bound on the total number of references was set to at least 200 references, employing a three-year citation window.[2] Additionally, the growth should be positive for the last five years, that is, every subsequent year should have more references than the year before. Finally, the last two years should show a growth that is either larger than the average growth in the previous year, or larger than the average growth the in the last two years of the all other remaining pairs.

Affiliation patterns

The second search strategy, affiliation patterns (AFP) characterises research fields by textual patterns for the affiliations (addresses) associated with publications. The list of patterns was created as follows. For each JSC, frequency lists of affiliation names were collected. Based on these lists, patterns such as ‘econ*’ for *Economics* or ‘fam*’ for *Family studies* were devised manually. Additionally, related JSCs were combined and used one affiliation pattern, resulting in 121 patterns for 243 JSCs.

Interdisciplinary connections are characterised as the co-occurrence of different patterns in the list of affiliations associated with a single paper, as in Gundersen and Ziliak (2004) where Department of Human Development and *Fam*ily Studies of the Iowa State University co-authored with the Department of *Econ*omics of the University of Kentucky.

Experience with the FFR strategy caused a reconsideration of the lower bound on the number of publications, notion of significant growth and the number of years in the analysis: a single year with at least 10 publications was used as the lower bound, the RTG was used to rank the most promising ones at the top, and we now use 11 years of data (1995–2005), because we focus on recent developments. The choice for 10 publications was based on the inspection of the distribution of the RTG, which appeared to resemble a normal distribution[3] without too many outliers when at least 10 publications were required. The inspection of the sorted list was stopped after inspecting 50 pairs.

Keyword sets

The third and final search strategy characterises research fields as sets of keywords (KWS). The keywords we use are the ‘keywords plus’, which are associated with many of the publications in the WoS (Garfield and Sher, 1993). For each JSC, keywords are selected based on publications published in the first two years (1995–1996). The objective of this selection is to catch the cognitive core of the fields, that is, the content which differentiates the fields. To accomplish this, the top 20% of most frequently used keywords is selected first. Then, to reduce overlap between sets, only keywords were kept which had at least 50% of their covered publications in one JSC. Finally, to reduce the processing requirements, we selected at most 500 of the keywords sorted on ascending coverage percentage and descending frequency. Based on our experience, 500 can be considered large enough. Additionally, we did not create sets for multidisciplinary JSCs as well as general, interdisciplinary JSCs, because early inspection revealed that such categories had too many words in common with too many other categories, and negatively influenced the set creation.

Inspection

We inspected selected pairs as follows. First, we dismissed pairs judged to be related *a priori*, usually pairs of JSCs with similar scope. Such judgements used objective criteria, such as the number of journals in common, but also background knowledge. Background knowledge was also used to dismiss ‘old convergence’, that is, pairs we knew to represent the result of convergence, but which took place some time ago. Next, the content of remaining pairs was inspected, with the objective of finding a limited number of subjects that would indicate the research focus of the convergence. Candidates would be dismissed if too many subjects were found or if the subjects were too unrelated. Finally, the candidates were given tentative names that reflected their research focus and passed on to experts.

Results

Application of the search strategies

In this section we present the application our search strategies and the inspection and validation of the results. The combined result of all strategies is given in Table 1.

Field-to-field references About 37,000 pairs contained at least one reference from the citing to the cited category, and about 5,900 pairs had at least 200 references in 15 years. Applying our growth requirements reduced this to 38 candidate pairs. Of these 38 pairs, those that had too many journals in common were dismissed, because we considered such pairs as representing well-known converged areas (e.g. nanotechnology). The final selection consisted of 11 pairs. Some of these pairs showed enough overlap to combine them in a single research area. Consequently, the 11 pairs were grouped into four converging research areas. In Table 1 these four areas are indicated with the FFR label in the strategy column.

Affiliation patterns About 5,200 pairs of patterns had at least one publication in common and almost 3,000 pairs had at least one year with 10 or more publications. Unfortunately, no research focus could be found in the top 50 pairs sorted on descending RTG. Instead, publications referred only research in one of the fields, or to research in fields that appeared not related to the affiliation, or contained a range of subjects instead of only limited number. For example, the pair *Opthalmology* (opth*) and *Neuroimaging* (neuroimag*) was a selected candidate, but the associated publications dealt with variety of (in our opinion) unrelated topics, such as ‘fatal insomnia’ and ‘kinase inhibitors’.

Keyword sets Over 7,500 pairs of sets had at least one publication in common and almost 3,000 pairs had at least one year with 10 or more publications. These pairs were sorted on descending RTG and the top 50 was inspected. After inspection, we selected 15 pairs. Again, some of these pairs again showed overlap and were grouped together. The final result consisted of 13 research areas. In Table 1 these areas are indicated with a KWS in the strategy column.

Expert verification The combined result from the FFR and KWS strategies consisted of 17 research areas, together with a description of their content. This result was presented to field experts, who were asked if they recognised the research area and if so, whether the area was the result of a convergence of other areas. Additionally, they were asked to verify or improve the (tentative) names given to the areas. The verification resulted in nine confirmed research areas, three areas judged too old, and five dismissed research areas, either because they were

Table 1. The research areas identified using the different strategies, together with their expert assessment and the candidate pairs grouped together into one area

Research areas	Strategy	#	Category 1	Category 2	Size
			Accepted		
Biomaterials	FFR	1	Chemistry, multidisciplinary	Engineering, biomedial	9,806
		2	Chemistry, multidisciplinary	Materials science, biomaterials	
		3	Chemistry, physical	Materials science, biomaterials	
Biosensors and biodevices	FFR	1	Biochemical research methods	Electrochemistry	2,982
ICT networks	FFR	1	Computer science, theory and methods	Telecommunications	8,474
Health aspects of nutritional components	KWS	1	Food science and technology	Endocrinology and metabolism	224
		2	Nutrition and dietetics	Gerontology	
Mild cognitive impairment and cerebrovascular disease	KWS	1	Peripheral vascular disease	Neuroscience	60
Neuro imaging and brain imaging	KWS	1	Radiology, nuclear medicine and medical imaging	Psychology, experimental	215
Polymers and the central nervous system	KWS	1	Polymer science	Neuroscience	75
Quantum information science	KWS	1	Polymer science	Engineering, electric and electronic	145
Surface physics and chemistry for biological and medical applications	KWS	1	Physics, applied	Dentistry, oral surgery and medicine	107
			Too old		
Bioinformatics	FFR	1	Biochemistry and molecular biology	Statistics and probability	
		2	Biotechnology and applied microbiology	Statistics and probability	
		3	Computer science, interdisciplinary applications	Genetics and heredity	
		4	Computer science, theory and methods	Biotechnology and applied microbiology	
		5	Genetics and heredity	Mathematics, interdisciplinary applications	
		6	Statistics and probability	Biotechnology and applied microbiology	
			Dismissed		
Genetic deficiencies and dermatological diseases in relation with neurological phenomena	KWS	1	Dermatology	Clinical neurology	54
Glasses for fibre lasers and amplifiers	KWS	1	Materials science, ceramics	Engineering, electrical and electronic	25
Simulation and modeling of nanotubes and hollow fibre membranes	KWS	1	Microbiology	Agriculture, soil science	35
		2	Chemistry analytical	Agriculture, soil science	
Socio-psychological aspects of socio-economic changes and circumstances	KWS	1	Developmental psychology	Economics	91
Soil microbiology and soil analytical chemistry	KWS	1	Mathematics	Engineering, chemical	86

Note: The overlap was calculated as the total number of (citing) publications published between 1995 and 2005

not recognised or were not the result of convergence. This result is reflected by the different parts ‘Accepted’, ‘Too old’, and ‘Rejected’ in Table 1.

Additional analyses

The differences in the overlap column of Table 1, which is representative for the size of the interdisciplinary area, are immediately apparent. In the case of the FFR strategy this was measured by the number of citing publications, while in case of the KWS strategy it was measured by the number of publications associated with both keyword sets. One can also note this difference in size by the names given to the research areas, which are shorter and more general for the larger areas, identified using the FFR strategy, while they are longer and more specific for the smaller KWS areas.

To further investigate this difference, Table 2 contains a summary of the distribution of the number of publications in the overlap, measured between 1995

Table 2. A summary of the size of the different research area characterisations, measured by the number of (citing) publications published between 1995 and 2005

Characterisation	Min	Median	Mean	Max
FFR	3,243	48,370	78,180	654,700
AFP	20	31,450	88,160	1,115,00
KWS	13	4,368	16,080	154,700

Table 3. The combined number of publications involved in the overlap between research areas

Strategy	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
FFR	965	1,196	1,302	1,568	1,999	2,474	3,082	4,010	6,616	8,664	11,385
KWS	22	37	59	55	79	92	115	133	154	182	330

and 2005. Table 2 also includes the numbers for the AFP strategy, because these numbers could hint at reasons for the failure of the AFP approach. Table 2 shows that the coarseness of the categories and the affiliation patterns are comparable, but that the spread of the size is wider in case of the affiliation patterns. Additionally, it shows that the size of a publication set covered by a keyword set is on average about 20% of that of a category, and about 18% of that of an affiliation pattern. However, the difference measured by the median is even larger, and then the size of the keyword sets is only respectively 9% and 13%.

Furthermore, for the FFR and KWS approach, Table 3 shows the size from year to year. These numbers are also shown in Figure 1, but normalised by setting the value in 1995 to 1.0. Figure 1 shows that, although the difference in absolute numbers of papers in Table 3 is large, the development of the growth has a comparable shape for both strategies.

A final additional analysis is given in Table 4, which shows the major science fields involved in the validated converging research areas. Table 4 was created by taking the JSCs involved in the research area characterisations, and aggregating those to their major science fields.[4] From Table 4 it is clear that the life sciences and the natural sciences dominate the results, while only one out of nine areas involved work from the social sciences and two out of nine involved work from engineering. Even more, the two fields involving work from engineering are never without JSCs that are also associated with the life sciences or the natural sciences.

Discussion

We presented three strategies for locating converging research using publication and citation data from the WoS database. Converging research was defined as the emergence of an interdisciplinary area from areas between which limited interdisciplinary connections existed before. This type of emerging research has not been studied extensively using publication data before, and it was not known beforehand what the different strategies could yield. It is therefore encouraging to see that our work yielded 26 converging pairs, which were grouped into 17research areas. Nine of these passed the judgement of experts, while three others were confirmed but dismissed as ‘too old’. It was found that a large size difference exists between the converged pairs resulting from the FFR strategy and the KWS strategy. However, the shape of the developments in the resulting areas was found to be comparable for both strategies. Finally, the results were mainly from the life sciences and the natural sciences.

It is difficult to judge which strategy was more fruitful: although more areas were identified using the KWS strategy (13 in total), more than half (eight) of these areas were considered too old or were otherwise dismissed. On the other hand, only four areas were identified using the FFR strategy, yet only one was dismissed as too old. At the same time it is not clear why the AFP strategy could not help to identify converging research: although patterns of fast-growing co-occurrences of different

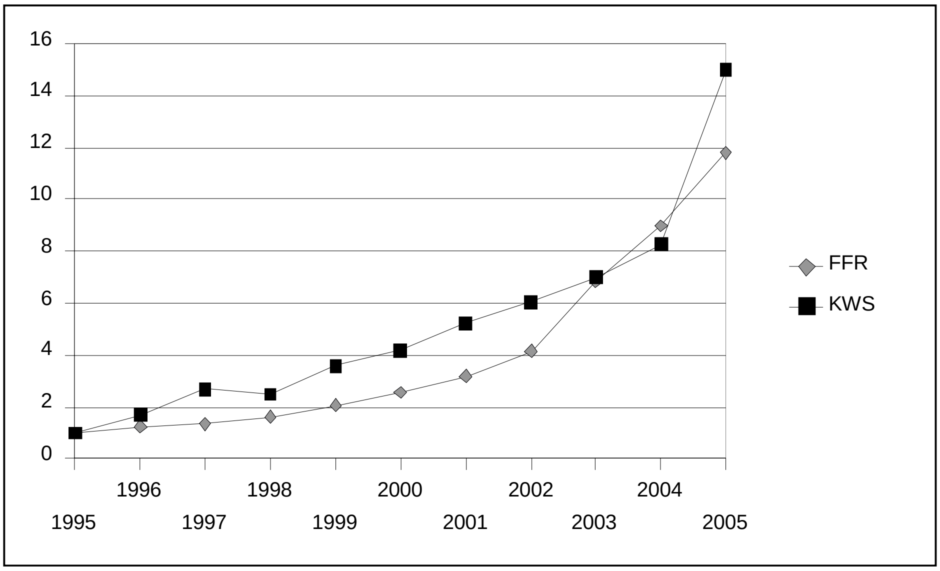

Figure 1. The combined number of publications involved in the overlap between research areas (Table 3) indexed to 1995

Table 4. Main fields of science involved in the detection of the selected research areas

Research area	Strategy	Main fields: Life	Natural	Engineering	Social
Biomaterials	FFR	Yes	Yes	No	No
Biosensors and biodevices	FFR	Yes	Yes	No	No
ICT networks	FFR	No	Yes	Yes	No
Health aspects of nutritional components	KWS	Yes	No	No	No
Mild cognitive impairments and cardiovascular diseases	KWS	Yes	No	No	No
Neuroimaging and brain imaging	KWS	Yes	No	No	Yes
Polymers and the central nervous system	KWS	Yes	Yes	No	No
Quantum information science	KWS	No	Yes	Yes	No
Surface physics and chemistry for biological and medical applications	KWS	Yes	Yes	No	No
Total involvement		**7**	**6**	**2**	**1**

affiliation patterns were found, none appeared to focus on the converging pairs, or on a limited number of subjects. Together with the distribution of the size of the overlap between affiliation patterns in Table 2, this may suggest that the current patterns are too multidisciplinary and may fail to provide enough *a priori* focus.

To give an example, the JSC *Crystallography* has a maximum overlap of 8% with other JSCs, but the affiliation pattern for *Crystallography* fails to reflect this: only about 25% of the papers are associated with only the JSC *Crystallography* and the rest with other JSCs. One could argue that this may indicate a multidisciplinary nature of the research in *Crystallography*, but it also appears not to provide a useful starting point to search for converging research. Perhaps this is for the same reasons for which we excluded the multidisciplinary JSCs in the KWS strategy, but it nevertheless deserves more attention in future research.

Review of the methodology and future research

An important advantage of our approach is that it has a quantitative basis. Therefore, it becomes possible to repeat and update the search with new data or fine-tuned parameters. Additionally, the application of a search strategy on the whole of science avoids a thematic bias and therefore improves the objectiveness of results. Other approaches, such as interviewing experts, may run the risk of a possibly biased view. On the other hand, experts may be able to signal developments sooner, since they may be part of specific developments, and are therefore not limited by the publication delays that are inherent to approaches based on peer-reviewed publication data. However, it is not an empirical fact that peers are always more up to date than our approach.

Some aspects in the strategies can be improved. Already during the project, the short-term focus provided by the three-year citation window used in the FFR approach was considered too short, given that interdisciplinary knowledge transfer takes time (Rinia *et al*, 2001, 2002). Also, a lower limit of 200 citations was high and possibly resulted in a bias toward large (biotechnological) fields. Improved versions of the strategies should select pairs on the basis of growth characteristics instead of (only) size characteristics. Also, the three strategies are (directly or indirectly) based upon the JSCs.

Although the JSCs have certainly been extremely convenient in the development of the strategies, alternative groupings could also be employed. For example, the use of smaller groups of journals based on journal citation patterns could improve both the FFR and the KWS strategies. The KWS strategy ‘freezes’ the keywords in the years 1995 and 1996. This was a conscious design choice, since this limited the use of the JSC on which it was based to only those years. However, this implies that knowledge developed after 1995 and 1996, which may use new and not included keywords, is disregarded and newer developments are not included. Another problem of the KWS strategy is due to homonymy and synonymy. For example, it is not possible to distinguish between ‘AD’ as an abbreviation for ‘Alzheimer’s disease’ or for ‘advertisement’.

Finally, the removal of the multidisciplinary categories was another design choice, because those categories made the distinction between other fields less clear, clouding the interdisciplinary developments. However, multidisciplinary JCSs can also be regarded as a bridge between fields and a basis for new interdisciplinary developments. Therefore, future research should focus on how to reintegrate them, or to use them as a source of information.

The current approach requires a qualitative selection process based on a judgement of resulting publications. Although this introduces a subjective element in the search, this also allows the process to include background knowledge which is difficult to express in quantitative terms alone. The use of background knowledge is witnessed by the small number of publications in some of the KWS pairs, which were chosen because both we and the experts recognised them to represent known research, without

The current approach requires a qualitative selection process based on a judgement of resulting publications, allowing the process to include background knowledge which is difficult to express in quantitative terms alone

covering all publications associated with that research. At the same time, we also dismissed pairs as too old using this kind of background knowledge. Additional tools could be used to try to encode such background knowledge and use that in the search and inspection. For example, Buter *et al* (2006) describes an interface between a (quantitative) cognitive science map, which encodes the content of a set of publications, and a qualitative, user-provided concept map that encodes background knowledge in the form of research topics.

Another extension which can be used in the inspection phase of the approach is to expand the number of publications by those which share an important part of the knowledge base (such as the best-cited publications). Such an extended set could create a more complete picture of the focus of the research in the convergence. To establish coherence of the (expanded) publication sets and aid the search for focus in the research, visualisation tools can help, such as cognitive maps (Buter and Noyons, 2001, 2002) or networks (Calero *et al*, 2006; Takeda and Kajikawa, 2009).

On the nature of converging research

We close this discussion with thoughts on the nature of converging research, combining theory and results described above. The original idea of converging research (Roco and Bainbridge, 2003) was more of a conceptual, high-level nature, rather than a concrete phenomenon. It was based on the idea that a unification of the sciences based on a ‘unity in nature’ (Roco and Bainbridge, 2003: 15) could enable new developments. This conceptual framework was put in more concrete terms in this article by defining converging research as emerging interdisciplinary research areas.

Also, there were no explicit or implicit restrictions put on the fields that could converge; and in principle any (exotic) combination was possible. However, many of the pairs found in Table 1 appear already related, even though they are different fields without shared journals or (initially without) shared keywords. For example, it appears unsurprising that *Multidisciplinary chemistry* may use results from *Materials science, biomaterials*; or that *Biochemical research methods* cites *Electrochemistry*. This also confirms earlier observations reported by Porter and Chubin (1985) who also found that the use of research results from distant fields is a rarity, using a limited set of journals and categories.

A possible reason for this is that such endeavours have an implicit risk for the researcher. So importing knowledge from a completely different domain may be less productive (Palmer, 1999). Additionally, if a researcher ventures into a new community, there is a need to learn and conform to existing vocabulary, descriptions of tools as well as problems. Acceptance and recognition by a new community may also be a boundary for a successful cross-over. Successful convergence therefore requires a mode of working that has a common language or understanding as well as a way to provide recognition of researchers from both originating disciplines. A ‘unity in nature’ is usually not enough.

At the same time, Palmer (1999) also notes that importing knowledge from other domains offers a better opportunity for knowledge-base development, and provides researchers with the opportunity to distinguish themselves by applying methodology to a new problem domain or (vice versa) to attack a problem with new methodology. This is also what we see in our results: science-born applications of tools on problems, where both tools and problems can be imported. For instance, the JSC *Computer science, theory and methods* imports problems from the JSC *Telecommunications* (related to networks), and applies its own methods in the converging research area. In the *Bioinformatics* converging research area, however, the *Genetics* and *Biochemistry* fields import tools from *Statistics*, for instance in order to apply these tools to expression data from micro-arrays.

Such developments result from the creative ‘probing’ efforts of researchers into new research directions, or caused by interdisciplinary developments related to tasks, processes, products, or use (Porter and Chubin, 1985). Those processes are part of normal science and are not necessarily indicative for Kuhnian paradigm shifts.

Concluding remarks

The results presented in this article show that converging research can successfully be identified using search strategies applied to publication and citation data. Although convergence was originally developed on a more conceptual level, the translation into detectable, scientific developments makes it an interesting evaluation tool, because it can highlight new interdisciplinary research developments with a novel character. These developments are interesting for at least two reasons: first, converging research areas are spontaneous, normal developments in science; and second, convergence occurs between fields that have some kind of conceptual distance between them. Especially the last point makes convergence a potentially valuable phenomenon, because combining knowledge domains requires creativity and effort, and is also not without risk. Therefore, to invest such effort and run such risks, potential benefits must appear valuable. At the same time, this also implies that most convergence was (and will be) found between fields that are not too distant from each other, since the effort to cross fields increases proportionally to the conceptual distance between them.

We conclude that, even though our strategies still require more development and the process as a whole would benefit from additional tools, the quantitative core of our approach could serve as the basis for a monitor that would signal new converging developments. Such a monitor could be a part of a policy instrument to identify areas where tools and problems are combined in new and creative ways, some of which may help to address current social, economic and scientific issues.

Acknowledgements

The authors of this article are grateful for the work done by our colleagues in the EMCOTEC project: Christien Enzing and Marc van Lieshout from the Netherlands Organisation for Applied Scientific Research (TNO), and Vera Grimm and Dirk Holtmannspötter from the Association of German Engineers (VDI). Also, we are indebted to the participants of the final EMCOTEC conference in Brussels (March 2008) who helped us shape our thoughts with their presentations and discussions. Finally, this work was done with support from the EU, as it was part of the 'New Emerging/Converging Clusters of Science and Technology' (EMCOTEC) project, contract SDPF-CT-2005-00002.

Notes

1. The WoS database available to the CWTS contains publication data from 1980 and onwards.
2. A citation window is the maximum number of years between the citing and the cited publication.
3. This was based on visual inspection and not verified statistically.
4. A CWTS taxonomy is being maintained which assigns JSCs to science fields and major fields.

References

Berg, J and R Wagner-Döbler 1996. A multidimensional analysis of scientific dynamics. *Scientometrics*, **35**(3), 321–346.

Besselink, M G, H C Van Santvoort, E Buskens, M A Boermeester, H van Goor, H M Timmerman *et al* (2008). Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. *Lancet*, **371**(9613), 651–659.

Budd, J and C D Hurt 1991. Superstring theory: information transfer in an emerging field. *Scientometrics*, **21**(1), 87–98.

Buter, R K and E C M Noyons 2001. Improving the functionality of interactive bibliometric science maps. *Scientometrics*, **51**(1), 55–68.

Buter, R K and E C M Noyons 2002. Using bibliometric maps to visualise term distribution in scientific papers. *Sixth International Conference on Information Visualisation (IV'02)*, 697–702.

Buter, R K, E C M Noyons, M van Mackelenbergh and T Laine 2006. Combining concept maps and bibliometric maps: first explorations. *Scientometrics*, **66**(2), 377–387.

Calero, C, R K Buter, C Cabello Valdés and E C M Noyons 2006. How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, **66**(2), 365–376.

De Solla Price, D 1981. The analysis of square matrices of scientometric transactions. *Scientometrics*, **3**(1), 55–63.

Garfield, E and I H Sher 1993. Keywords plus algorithmic derivative indexing. *Journal of the American Institute for Information Science*, **44**, 298–299.

Goffman, W 1971. A mathematical method for analyzing the growth of a scientific discipline. *Journal of the Association for Computing Machinery*, **18**(2), 1672–1673.

Gundersen, C and J P Ziliak 2004. Poverty and macroeconomic performance across space, race, and family structure. *Demography*, **41**(1), 61–86.

Haniel, A 2000. Organs at any price? On the question of alternatives to postmortem organ donation (examining ethical aspects of human organ cultivation and cloning). *Zeitschrift für Evangelische Ethik*, **44**(4), 269–284.

Hinze, S 1994. Bibliographical cartography of an emerging interdisciplinary discipline: the case of bioelectronics. *Scientometrics*, **29**(3), 353–376.

Hammer, C, R Linke, F Wagner and M Diefenbeck 1998. Organs from animals for man. *International Archives of Allergy and Immunology*, **116**(1), 5–21.

May, K O 1966. Quantitative growth of the mathematical literature. *Science*, **154**, 1672–1673.

Morillo, F, M Bordons and I Gómez 2003. Interdisciplinarity in science: a tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, **54**(13), 1237–1249.

Morris, S A 2005. Manifestation of emerging specialties in journal literature: a growth model of papers, references, exemplars, bibliographic coupling, cocitation and clustering coefficient distribution. *Journal of the American Society for Information Science and Technology*, **56**(12), 1250–1273.

Nordmann, A 2004. *Converging Technologies for a Diverse Europe*. Brussels.

Palmer, C L 1999. Structures and strategies of interdisciplinary science. *Journal of the American Society for Information Science*, **50**(3), 242–253.

Porter, A and D Chubin 1985. An indicator of cross-disciplinary research. *Scientometrics*, **8**(3), 161–176.

Rinia, E, T N van Leeuwen, E Bruins, H van Vuren and A F J van Raan 2001. Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, **51**(1), 293–309.

Rinia, E, T N van Leeuwen, E Bruins, H van Vuren and A F J van Raan 2002. Measuring knowledge transfer between fields of science. *Scientometrics*, **54**(3), 347–362.

Roco, M C and W S Bainbridge eds 2003. *Converging Technologies for Improving Human Performance*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Ródenas-Torralba, E, A Morales-Rubio and M de la Guardia 2006. Scientometric picture of the evolution of the literature of automation in spectroscopy and its current state. *Spectroscopy Letters*, **39**, 513–432.

Schummer, J 2004. Multidisciplinarity, interdisciplinarity and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, **59**(3), 425–465.

Shibata, N, Y Kajikawa, Y Takeda and K Matsushima 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, **28**(11), 758–775.

Tabah, A N 1999. Literature dynamics: studies on growth, diffusion and epidemics. In *Annual Review of Information Science and Technology*, ed. M E Williams, **34**, 249–286.

Takeda, Y and Y Kajikawa 2009. Optics: a bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, **78**(3), 543–558.

Van Raan, A F J 2008. Bibliometric statistical properties of the 100 largest European research universities: prevalent scaling rules in the science system. *Journal of the American Society for Information Science and Technology*, **59**(3), 461–475.