Capturing new developments in an emerging technology:
An updated search strategy for identifying nanotechnology research outputs

Sanjay K. Arora[a], Alan L. Porter [a], Jan Youtie[b], Philip Shapira[a,c,*]

[a] School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA

[b] Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA

[c] Manchester Institute of Innovation Research, Manchester Business School, University of Manchester, M13 9PL, UK

[*] Corresponding author: *Email address:* pshapira@mbs.ac.uk

May 2012 (Working Paper – Pre-Print)

Abstract

Bibliometric analysis of publication metadata is an important tool for investigating emerging fields of technology. However, the application of field definitions to define an emerging technology is complicated by ongoing and at times rapid change in the underlying technology itself. There is limited prior work on adapting the bibliometric definitions of emerging technologies as these technologies change over time. The paper addresses this gap. We draw on the example of the modular keyword nanotechnology search strategy developed at Georgia Institute of Technology (Georgia Tech) in 2006. This search approach has seen extensive use in analyzing emerging trends in nanotechnology research and innovation. Yet with the growth of the nanotechnology field, novel materials, particles, technologies, and tools have appeared. We report on the process and results of extensively reviewing and revising the Georgia Tech nanotechnology search strategy. By employing structured text-mining software to profile keyword terms, and by soliciting input from domain experts, we identify new nanotechnology-related keywords. We retroactively apply the revised evolutionary lexical query to twenty years of publication data to produce a powerful and rich panel dataset. Our findings indicate that the new search approach offers an incremental improvement over the original strategy in terms of recall and precision. Additionally, the new strategy reveals several emerging cited subject categories particularly in the biomedical sciences, suggesting a further extension of the nanotechnology knowledge domain. The implications of the work for applying bibliometric definitions to emerging technologies are discussed.

Key words:

Nanotechnology; Bibliometric; Publications; Search Strategy; Cited Subject Categories.

MSC Classification:

91

JEL Classes:

C89; O30

Introduction

Nanotechnology is a broad and growing domain that involves the understanding and engineering of matter in the nanoscale dimensional range of 1 to 100 nanometers (nm). Distinctive and novel physical, chemical, and biological properties and features result from the manipulation of nanoscale particles, materials and systems (President's Council of Advisors on Science and Technology, 2010). Research in nanotechnology spans the spectrum of established and emergent scientific and technological disciplines including physics, chemistry, material science, engineering and biotechnology

The particular characteristics of nanotechnology and the array of disciplines involved in its research and development present several challenges for the creation of bibliometric definitions of the field. The size-specific criterion of what constitutes nanotechnology cannot be reliably used by itself as a mechanism to distinguish scholarly literature in the field (NSTC, 2007). Employing subject category classifications does little to help demarcate the borders of nanotechnology. Furthermore, journals with nanotechnology or "nano" in the publication name may no longer indicate a propensity to exclusively focus on nanotechnology (Grieneisen 2010). Nevertheless, capturing the growth and development as well as the other more nuanced qualities of nanotechnology scientific output is critical to understanding the evolution of the domain, the emergence of new technological and commercial opportunities, and potential societal and risk implications.

For several years, the Nanotechnology Research and Innovation Systems Assessment group at Georgia Institute of Technology (Georgia Tech) has been tracking the development of nanotechnology research and innovation. For bibliometric research on nanotechnology, a key tool has been the development of an encompassing definition of the nanotechnology domain. We initiated this effort in 2005, with calibration and analysis of findings appearing in the period 2006 onwards. Our nanotechnology search approach comprised a modular keyword search strategy with a two-step inclusion and exclusion process. The first full application of the search approach identified more than 406,000 nanotechnology Web of Science papers and over 53,000 MicroPatent and INPADOC patent records published between 1990 and mid-2006 (for full details of the search approach and initial results, see Porter et al., 2008). With the worldwide expansion of funding and activity in nanotechnology in recent years, the number of records captured by further runs of the search approach grew. For example, by mid-2011, our nanotechnology search approach was identifying more than 820,000 Web of Science papers published since 1990.

We have used this search approach in studies that have examined a series of questions and topics related to nanotechnology research and innovation and its implications, including identifying trajectories of nanotechnology publications and patents (Youtie et al., 2008), funding sponsorship of nanotechnology research (Shapira and Wang, 2010), the development of active nanotechnologies (Subramanian et al., 2010), national and regional nanotechnology emergence (Shapira and Youtie, 2008), and nanotechnology's interdisciplinary linkages (Porter and Youtie, 2009). The approach performed robustly when compared with other nanotechnology search strategies (Huang et al., 2010) and findings based on the approach have been referenced not only by other researchers but also in policy documents (for example, see PCAST 2012).

While the original search approach is comprehensive, as the elapsed time from the original definition point increases and as the science and technology of nanotechnology evolves, questions arise as to whether the search is capturing new developments and topics. For instance, graphene, a nanoscale material comprised of a single layer of carbon atoms that was identified and characterized less than ten years ago, has seen rapid growth in scientific and patenting activity recently and was the subject of the 2010 Nobel Prize in Physics. Yet, the keyword "graphene" was not explicitly included in our initial nanotechnology search strategy. Such an omission would not be detrimental if graphene articles were captured via another term included in the original search query. However, if this were not the case, it would suggest the need to update the approach not only to capture this new topic but also to investigate other new topics and to verify the overall performance of the search.

This illustration highlights a broader underlying question. Although a search approach may have performed well historically, inevitably it will begin to lose both precision and recall over time and will need to be reviewed. As a scientific domain evolves over a period of years, when is it appropriate to update a bibliometric search strategy? In a domain as sizeable as nanotechnology this is a critical issue, as the updating process will likely require significant investments in time and resources to implement. This research question is at the heart of this paper, which develops an approach to updating our nanotechnology search terminology to reflect the evolving nature of the field.

The paper is organized in the following manner. First, we review the original nanotechnology search approach and associated literature in the context of approaches put forward by other researchers for delineating the nanotechnology domain. Here, we characterize the complexities of information retrieval in general and focus more specifically on tradeoffs between precision and recall in large scale bibliometric analysis. The ensuing section presents our updated methodology for identifying research outputs in nanotechnology. We then test the performance of the updated search strategy, and present the results. Finally, we conclude with a discussion of implications and future areas of application.

Context and Literature Review

The initial nanotechnology search strategy was devised with three principal aims in mind (see Porter et al., 2008). First, each search term component had to add value to the overall search by attracting a non-trivial number of unique publication records. Second, the search approach had to be relatively uncomplicated yet also comprehensible so to domain experts. We did successfully receive feedback from experts in the nanotechnology domain, including researchers in academic departments, interdisciplinary research centers, industry and government. Finally, the search strategy had to allow for the addition, removal, or modification of terms as the field evolved over time.

An important consideration in the formulation of the initial search was the inherent tradeoff between recall and precision. High recall signifies that a search query captures most, if not all, of the relevant records that would be identified under the most optimistic scenario (i.e., if the query was close to perfect in identifying all germane nanotechnology records). Precision, on the other hand, measures the number of truly relevant records returned by the query. A high degree of precision indicates that there is a limited amount of noise – or few irrelevant records – in the resulting dataset. Information scientists typically view the association between recall and precision as inversely related: high recall can only be attained at the expense of lower precision (Buckland and Gey, 1994). Our nanotechnology search approach sought to optimize between the extremes of high recall and high precision. We sought to capture a broad array of the nanotechnology literature (thus maximizing recall) while concomitantly avoiding certain keywords that produced too much noise. The initial strategy, therefore, excluded certain frequently occurring bio-oriented terms such as DNA, RNA, and biochip: while such terms are evident in nanotechnology research, they are far more commonly found in the wider life-sciences literature and to retain them would significantly reduce precision. We also excluded other specific keywords such as spintronic and molecular beacon as a result of our own evaluation and after expert review raised concerns regarding the extent to which these keywords would generate a high level of irrelevant results (Porter et al. (2008).

The initial search strategy consists of two steps. The first step applies a set of eight modular components ranging from the broadly encompassing query, “nano*”, to more granular queries considering nano-relevant applications (e.g. molecular wiring), sub-fields of nanotechnology (e.g. bionano*), and instrumentation and techniques for producing nano-related research (e.g. certain types of microscopy and lithography). Many of these individual modular components include terms that are contingent on other keywords being present. For example, all microscopy terms must be attendant with at least one other keyword from the “molecular environment”. The molecular environment in this case reflects such keywords as monolayer, film, and copolymer. For other terms that were clearly nanotechnology related, we included without molecular environment qualifiers. The contingency approach produces a record set with a higher degree of both precision and recall than would otherwise be possible if either the contingency or instrumentation terms were omitted. An eighth modular component consists not of topical search terms but rather publication sources. This accommodation accounts for articles published in nanotechnology-oriented journals that may not explicitly contain keywords found in the first seven modular query components.

The second step of the initial search strategy involves an exclusion process. This *removes* publication records captured by the first stage search where such exclusions will improve precision. We identified a set of about 40 “exclusion terms” which reflect measurements at the nanotechnology scale (to remove records that only reference a nanoscale measurement but have no other indications of nanotechnology content) and other spurious derivatives of the all-inclusive nano* query. Examples of exclusion terms include n*plankton, nanoalga*, nanobacteri*, nano2 (the chemical formula for sodium nitrate), nanometer*, and nanosecond* where such terms occurred singly in a record.

Other researchers have also developed search strategies to define nanotechnology publications. Huang et al. (2010) reviewed several of these approaches (including our initial nanotechnology search) and classified them into four main groups: lexical queries, evolutionary lexical queries, citation analyses, and publications listed in core nanotechnology journals. A lexical query relies on expert advice for keyword identification. Although relatively straightforward to implement, the reliability of lexical searches depends on the proficiency of the experts consulted. Our initial nanotechnology search was a lexical search which drew on a range of experts and an iterative validation procedure for candidate keyword identification prior to searching. An evolutionary lexical query employs semi-automated search term identification processes to discover trending keywords. Experts then offer their recommendations from a candidate list of keywords, thus minimizing researcher bias. While *in practice* the flow of information in question most likely dictates how experts discard or isolate certain keywords for inclusion or exclusion in the search strategy, it is also possible for new keywords to change the ways in which experts view the field. To develop their nanotechnology search strategy, Mogoutov and Kahane (2007) engage experts in the latter stages of their automated lexical query process, combining a static nano* query with an auto-generated list of subject discipline-specific keywords.

Huang et al. (2010) characterize a citation based search strategy as one that relies on a core set of literature in order to identify articles that cite the core. The exact "parameters" of the algorithm are defined and bound by the authors implementing the strategy, and therefore, this approach does not require expert input. A weakness with citation analysis, however, is in its portability and replicability (Mogoutov and Kahane, 2007). Researchers without a full suite of publication metadata cannot replicate a citation based domain definition of nanotechnology. Therefore, computation and licensing costs are both salient when considering whether citation analysis is a feasible alternative to lexical (keyword) querying. Zitt and Bassecoulard (2006) employ citation networks to expand their corpus of nanotechnology publications, beginning with a seed set of nanotechnology literature identified through a series of modular queries. By first identifying a set of core literature that the seed cites, the authors mark other articles that also cite the core. Once the core set of articles is identified (a critical first step), citation analysis is automated and relies on a handful of configurable parameters to keep only those articles that have a significant citing relationship with the core.

The final category of search strategies bases its methodology on a set of dedicated nanotechnology journals, thereby including all articles in that set of journals. Leydesdorff and Zhou (2007) offer a methodology that begins with a core set of six nanotechnology journals and, through citation and network analysis (using betweenness centrality), expands that core set to ten journals. A journal is a "core" publication if it contains "nano" in its title. Huang et al. (2010) note that while, in theory, precision should be relatively high with this method, recall suffers because nanotechnology research is published extensively outside the scope of the limited set of dedicated nanotechnology journals.

When these contrasting search approaches are tested and compared, our initial search strategy (Porter et al., 2008) performs well. Huang et al. (2010) examined our approach and its results along with five other leading nanotechnology search strategies. Porter et al. (2008) provides the second highest number of records (behind Mogoutov and Kahane, 2007) and offers a similar subject discipline composition to four of the five strategies (not including Leydesdorff, 2008). Cunningham and Porter (2011) provide a separate assessment of the Porter et al. (2008) approach by comparing the initial search definition with a series of auto-generated queries produced by machine learning algorithms. Machine learning offers a way to assess efficiency performance by determining whether there is an alternative, more parsimonious approach to identifying the set of articles in a search. The authors conclude that while some new terms could be added (e.g. graphene and epitaxy) and a few removed, the Porter et al. (2008) approach as a whole demonstrates high robustness.

As of May 2011, at least 55 articles (using the Web of Knowledge, all databases) and as many as 148 publications (using Google Scholar) cite the original Porter et al. (2008) paper. (These citations include subsequent papers by one or more of the original authors.) Many of the scholarly articles explore empirical research work that either relies directly on or is (indirectly) informed by this nanotechnology search strategy. One stream of research examines the spatial distribution of general nanotechnology activity along national or regional lines (e.g. López Cadenas et al. 2011; Shapira and Youtie, 2008), often in comparative terms (e.g. Kay and Shapira, 2009), while another group of articles focuses on specific emerging nanotechnologies (e.g. Subramanian et al., 2010; Guo et al., 2010). A few studies use the approach to assess environmental, safety, and health risks (EHS) of nanotechnology (e.g. Youtie et al, 2011); other work focuses on the nature of scientific collaboration (Pei and Porter, 2011), funding patterns at the institutional level and across national borders (Shapira and Wang, 2010), and the social dimensions of technology governance (Ho et al., 2010). Together, these citing references reflect a diverse body of subject disciplines, including Science and Technology Studies, Materials Science, Chemistry, Business & Economics, Computer Science, Engineering, Physics, and Public, Environmental & Occupational Health. As the social science literature grows around nanotechnology (Shapira et al, 2010), we see the need to ensure that the domain definition based on scientific output reflects the current, as well as historical, trajectory of the field.

None of the aforementioned search strategies for characterizing nanotechnology have published subsequent modifications to take into consideration the changing nature of the field. This paper addresses this gap. It presents a methodology for modifying the initial search to account for changes in nanotechnology and demonstrates the contribution of the lexical modifications to the query and resulting database. In addition to ensuring an updated an optimal balance between recall and precision in search results, this updating process also allows us to identify ways in which the nanotechnology domain has grown and evolved in the more than five years since the initial search strategy was devised. After presenting the methodology for updating the search, we thus discuss the new results produced and their interpretation and implications.

Methodology

As indicated above, the Porter et al. (2008) search strategy produced in 2005-2006, is a lexical approach that draws on expert opinion for keyword identification. Our second version of the search strategy, developed in the period 2011 to early 2012, can best be characterized as an *evolutionary* lexical query, which employs feedback channels between the keyword identification process and elicitation of expert opinion.

The methodology for revising the second version of the search strategy began with the collection of high-frequency nanotechnology keywords from a subset of records (those published in 2009 and those keywords matching “nano*”). We used VantagePoint – a software application for structured text mining and analysis (see www.theVantagePoint.com). Approximately 1,100 ensuing high-frequency key terms were identified and manually sorted into three categories (Figure 1). The first “accounted for” group contained keywords already included in the legacy search strategy. The second “ignore” group consisted of words that are apparently unrelated to nanotechnology (e.g. publishing artifacts). The third “investigate” group signaled certain keywords not included in the original search strategy and which thus warranted additional analysis.

[INSERT FIGURE 1 ABOUT HERE]

At this stage, we devised a simple noise ratio that captured the extent to which a given keyword related to the field of nanotechnology. We used the same nanotechnology subset of records as well as a 40,000 record random sample of Web of Science 2009 publications. By using the native search capabilities of VantagePoint, we computed the following noise ratio value for each keyword in the “investigate” group:

noise ratio = (number of hits in random sample / 40,000) ÷

(number of hits in nanotechnology subset / total number of records in all 2009 nanotechnology records)

Eliminating keywords with values below a noise threshold of 20% produced a list of 75 candidate search terms, some of which could be combined because of obvious lexical similarities or through lemmatization. A more parsimonious threshold of 8.5% yielded ten candidate keywords. Using the 75 keywords would lead to a broader definition of nanotechnology and thus require some additional screening to maximize precision. However, because of the evolutionary nature of our approach, however, we decided to pursue only those keywords that would result in minimum additional noise. This top-trending analysis of frequently appearing terms produced a handful of new terms for the final search strategy. Appendix Table 1A presents a sample set of candidate keywords and their noise ratios.

In addition to the semi-automated approach of identifying top-trending keywords, we also evaluated noise ratios for potential key terms identified through sources, including popular press coverage of nanotechnology, interaction with scientists, and a review of notes from the Porter et al. (2008) study. Formal feedback from experts was an essential input to refining the query terms. We began with individual meetings at Georgia Tech with three nanotechnology specialists: a research scientist, a research engineer, and a doctoral student (offering diversity in expertise). During our conversations, we gained valuable feedback on new as well as old keywords. Additionally, we piloted a brief survey that asked several questions with respect to the scope and accuracy of our modular approach. We subsequently sent this questionnaire to 67 contacts in the US and internationally, which included many research scientists and professors, several industry and government practitioners, and one representative from each of the fourteen US NNIN (National Nanotechnology Infrastructure Network) centers. We received seventeen responses, with twelve completing the survey, resulting in a final response rate of about 18%. This response rate is common for such voluntary surveys. Most important for us was the quality and detail of the responses and the range of, as well as the level of consensus about, suggested key nanotechnology terms. Additionally, prior to the survey, we held several personal meetings with nano-scientists and engineers from various subfields of nanotechnology, which typically lasted for about an hour. These in-depth meetings produced a about 100 possible keywords in total, some of which overlapped with the terms found in the top-trending search process. Consequently, we applied the noise ratio to 87 of these unique search term combinations and kept only those keywords that met or exceeded our 8.5% noise threshold requirement. Then, after incorporating these new changes into the questionnaire, we sent the surveys and subsequently received responses that either validated or contested some of the additions. For instance, after adding “contact angle,” we received sufficient concern to remove the search term in the final iteration. Several new terms were identified as possible additions from the written survey responses, including ferrofluids, core-shell, magnetic force microscopy, and MFM. We then consulted with expert contacts at Georgia Tech to validate which of these keywords would be most advantageous to pursue.

An example of the value of this interactive process was the especially fruitful communication held with a scientist and National Nanotechnology Infrastructure Network site manager at Georgia Tech. This interaction revealed the need to amend the legacy microscopy and spectroscopy search terms. We had noticed some degree of noise associated with keywords such as scanning electron microscopy used in the initial search approach. During an in-person meeting, we provided a list of thirty article titles identified by this particular modular query; the articles were then categorized with the scientists as “nanotechnology” (13), “not nanotechnology” (7), and “possibly nanotechnology” (10). On investigation of the latter two groups, we recognized that some additional contingency terms (such as polymer, material, and molecule) which related to the inclusive molecular environment were contributing to sixteen of the seventeen suspect search hits. Therefore, in order to reduce noise, we revised the microscopy and spectroscopy query to be contingent on terms from the more restrictive molecular environment terms. As a result of this modification, we expect precision to increase dramatically, though partially at the expense of recall. By and large, however, we mitigate the reduction in contingency terms with an expansion of several new specific microscopy and spectroscopy keywords.

Tables 1 and 2 present the final search strategy of the second version of our nanotechnology search. Additions to the initial definition are emphasized in bold underlined text. With the exception of the eighth query component that focuses on nano-related publications of interest, the modular query is deployed against the title, abstract, and author keywords of a scientific article (using the "TS" qualifier in Web of Science). Some keywords contain an asterisk, which is used as a "wildcard" to designate other versions or spellings. For the first query component (nano*), we considered variations matching a*nano*, b*nano*, c*nano*, etc., but decided against such an approach due to the pervasiveness of many non-nanotechnology related terms corresponding to that pattern (e.g. allopregnanolone, mannanoligosaccharide, nonanoate, perfluorononanoic, and subnanomolar).

[INSERT TABLE 1 ABOUT HERE]

[INSERT TABLE 2 ABOUT HERE]

As in our initial search strategy, the second version uses a list of exclusion terms to remove unwanted and out-of-scope records (see Table 3). While some exclusion terms, if found in a given record, result in the removal of that record from the dataset *sine qua non*, other exclusion terms, particularly those related to measurements, result in the removal of records only if the record does not include another nano-related keyword. To the list of original (Porter et al., 2008) exclusion terms, we added spelling variants of measurements at the nanoscale, "nanosatellite", as well as approximately 270 organism names beginning with nano*, as identified by Grienheisen and Zhang (2011).

[INSERT TABLE 3 ABOUT HERE]

Results

We present the results in three subsections: a comparison of the performance of the initial and second search strategies; a brief look at national trends; and a detailed analysis of emerging subject categories and cited subject categories in the corpus of nanotechnology publications.

Performance

A comparison of the result sets returned from the initial and second search queries reveals a significant overlap in the number of records identified in any given year (Table 4). On a year-to-year basis, the number of common articles ranges from a low of 78% in 1990 to a high of 94% in 2010. At first glance, this finding suggests that the initial and second queries converge over time with respect to their projected domain definitions of nanotechnology. On closer inspection, however, we attribute this trend to the lower use of the "nano" prefix – vis-à-vis the other sub queries combined – in article topics in the early years of the domain in the 1990s. Whereas records identified by "nano*" generate less than 10% of the total number of retrieved records in 1990, this share increases to 76% in 2010. Thus, the keyword changes outlined in Table 1 have a greater impact on the search strategy in earlier years than on later years. In addition, the effect of exclusion terms on publication year totals indicates that records matching nano* in the 1990s are less likely to concern nanotechnology, per our domain definition, as publications from the 2000s; that is, the nano* prefix tends to capture more papers not relevant to nanotechnology in the 1990s than in the 2000s. Taken together, these findings suggest that over time researchers have formulated, or at least arrived at, an increasingly shared understanding of what is nanotechnology and that articles involving only nanoscale measurements or other non-relevant nano* terms represent a small and decreasing proportion of the expansion of nanotechnology publishing in recent years.

[INSERT TABLE 4 ABOUT HERE]

The initial and second version of our search approaches confirm (as found by other researchers) that there has been a marked increase in the total number of nanotechnology publications published annually (see Table 4). Our new search strategy identifies about 760,000 Web of Science nanotechnology papers published between 1990 and 2010; of these, some 2,400 were published in 1990, with over 23,000 published in 2010. Most years in the twenty year window saw double digit annual percent increases in nanotechnology publications.

National Trends

Within the overall growth in the production of nanotechnology papers, there are significant country-level differences, and also developments in the lead set of countries driving growth in nanotechnology outputs. While we have data for all countries where there are authors involved in nanotechnology publication activities, to focus the discussion, we present here results in two-year increments for the five most prolific producers across our twenty year time horizon (see Figure 2). Although both the US and China initiated national nanotechnology initiatives at about the same time in the early 2000s (Shapira and Wang, 2010), the US was the world's leading producer of nanotechnology publications for much of this decade. However, our search results confirm that the US has recently been out-produced in absolute terms by China, which now holds the global frontrunner position with over 20,000 publications in 2010. Germany, Japan, and

South Korea comprise the next set of producers by absolute size, with all three of these countries seeing steady year-over-year percent increases in output over the last decade.

[INSERT FIGURE 2 ABOUT HERE]

Emerging Research Areas

At this stage, we turn to characterizing the subject categories and cited subject categories of nanotechnology records as identified by the updated search query. Subject categories are based on classifications of journals used in Web of Science, drawing on the science mapping method developed by Leydesdorff, et al., 2012). Table 5 presents the top twenty subject categories in 2010 and compares how these rankings have changed since 2000 and 2005. Many of the relative rankings remain the same despite the ten year time period. For example, “Materials Science, Multidisciplinary”, “Physics, Applied”, “Chemistry, Physical”, and “Chemistry, Multidisciplinary” sustain consistent standing in the top five subject categories. However, two noticeable trends materialize. First, “Nanoscience & Nanotechnology”, introduced into Web of Science (WOS) in 2005, not surprisingly reflects rapid growth; as of December, 2011, 27 journal titles in the WOS Science Citation Index (SCI) and 66 journals in SCI Expanded belong to this subject category. Secondly, the rise of certain applied, cross-disciplinary subject categories, such as “Electrochemistry” and “Materials Science, Biomaterials”, at the expense of more single disciplinary subject categories, such as “Physics, Atomic, Molecular & Chemical” and “Engineering, Electrical & Electronic”, may signal that nanotechnology research is indeed becoming more applied as novel application areas leverage previous advancements in basic research at the molecular and atomic levels.

[INSERT TABLE 5 ABOUT HERE]

These observations should be interpreted with caution. Substantiating the factors behind changes in WOS subject categories rankings is tenuous with the data at hand. Subject categories are applied at the journal level, and all articles in a publication title inherit these classifications accordingly. It is certainly plausible that articles in a journal may not align accurately with the given subject category – or that the addition of new journals in a particular subject area skews the number of publication records in one sample time frame vis-à-vis another.

To better understand the nuances of subject categories as indicators of the development of nanotechnology as a whole, we turn to *cited* subject categories. By definition, the interdisciplinary (or multidisciplinary) nature of nanotechnology pulls on intellectual output from a variety of subject areas. Cited subject categories, derived from cited references, are likely to reflect a varied and nuanced proxy of knowledge links among discrete, disciplinary areas. Using VantagePoint, we capture journal citations then apply a thesaurus to obtain the corresponding cited subject categories. This approach gives us a proxy of the "research program", as initially described by Lakatos (1978). Lakatos defines a research program as consisting of a hard core of assumptions and a protective belt, which shapes and advances problem shifts. It is in the protective belt that we seek to explore nanotechnology's most recent problem shifts from the perspective of cited subject categories.

Our aim is to identify cited subject categories that exist at the periphery of the research program, in the protective belt, and that exhibit excessive empirical content. To accomplish this objective, we pare down the list of cited subject categories to include only those areas that have changed significantly in the three-year sample timeframe (i.e., in 2000, 2005, and 2010). In particular, we compare the rank order of cited subject categories from one period to the next, excluding those cited subject categories that experience a variation of less than (positive) four.

For instance, to isolate emerging cited subject categories in 2005, we compute the rank order of cited subject categories in 2000 and 2005 and then subtract rank values in 2000 from those in 2005. We also 1) eliminate "Nanoscience & Nanotechnology" due to its recent inclusion into the Web of Science typology and because of its role as an all-encompassing cited subject category, and 2) ignore cited subject categories with fewer than 500 total citations in the three year sample. To visualize the progressivity of the research program, we present network maps of the cited subject categories in 2005 versus 2010. The maps apply one additional filter to enable better visualization of results. We remove edges symbolizing fewer than 25 subject area co-citation occurrences for the 2005 data and remove edges representing fewer than 200 co-citation occurrences in 2010. All in all, the network maps portray emerging cited subject categories as nodes, with heavier edge weights indicating increased levels of co-citation occurrences. In other words, the network maps illustrate a subset of up-and-coming cited subject categories that are often co-cited within the corpus of nanotechnology publications.

The map of emerging cited subject categories in 2005 (see Figure 3), by meta-discipline, depicts a strong presence of subject categories related to biomedical sciences, which constitutes 19 of the 37 emergent cited subject categories. Rafols et al. (2010) have undertaken factor analysis of the subject category cross citation matrix for a target year (2007) of Web of Science publications to group them into macro-disciplines, and, in turn, meta-disciplines. Here we use four meta-disciplines as defined by Rafols and colleagues. From this, we see that physical sciences and environmental sciences contribute ten and eight cited subject categories, respectively. In 2010, the map depicts an even greater presence of cited subject categories in the biomedical sciences, which encompasses 25 out of 40 nodes (see Figure 4). The physical sciences and environmental sciences each maintain seven and eight emerging cited subject categories, respectively. It is worthwhile to note that many of the most highly cited subject categories such as “Materials Science, Multidisciplinary”, “Physics, Applied”, and “Physics, Condensed Matter” are not represented in the analysis because their positions in the relative rank order of cited subject categories have not changed much since 2000. Thus, this analysis focuses only on potential emerging areas of nanotechnology in recent years.

[INSERT FIGURE 3 ABOUT HERE]

[INSERT FIGURE 4 ABOUT HERE]

The network diagrams provide us with a summary level overview of how different up-and-coming subject categories align and connect; however, the visualizations do not confer precise indicators of importance and weight. Consequently, we turn to two measures of interest, number of citations to a particular subject category and eigenvector weighted centrality (see Table 6). Whereas number of citations reveals the number of references to articles in the emergent subject category, weighted eigenvector centrality offers a more nuanced measure that considers both the presence of ties to other nodes (i.e., subject categories) in the network as well as the importance of adjacent node weights (Newman, 2004). Again, we emphasize that edge weights equal the number of times one subject category has been cited along with another subject category within the same article in the corpus. In general, weighted eigenvector scores correspond to the largest eigenvalue of the symmetric weighted adjacency matrix (Bonacich, 2007). The eigenvector corresponding to the maximum eigenvalue contains only non-negative values, which in turn represent global centrality scores for each node (i.e., cited subject category) in the network (Runhau, 2000).

[INSERT TABLE 6 ABOUT HERE]

Ranked by weighted eigenvector centrality score, six of the top ten emerging cited subject categories in 2005 belong to physical and environmental sciences; that is, even though the 2005 network (Figure 3) contains 19 subject categories in the biomedical sciences, only four of these disciplines are ranked in the top ten by weighted eigenvector score. “Engineering, Chemical” attains the most citations overall (7,299) and the highest weighted eigenvector score (1.00), followed by “Environmental Sciences”, “Engineering, Environmental”, “Biotechnology & Applied Microbiology”, and so on. A brief comparison of the 2005 network diagram and the top ten cited subject categories ranked by eigenvector centrality exposes a cluster of central, emerging cited subject areas in the eastern sphere of the map. Co-citations are strong across adjacent nodes in this boundary area, suggesting a high degree of interdisciplinary engagement.

Using the same framework for investigation, the 2010 network, in conjunction with the top ten cited subject categories, implies that progressive problem shifts are becoming increasingly abundant in the biomedical arena. Furthermore, unlike the diagram for 2005, the locus of most emerging cited subject disciplines does not fall in the eastern sphere of the map. For instance, several of the cited subject categories, such as “Pharmacology & Pharmacy”, “Oncology”, and “Medical Laboratory Equipment”, are deeply embedded within the biomedical sciences portion of the map, suggesting that these emerging cited subject categories in nanotechnology are becoming more influential as time passes. Indeed, ranked by weighted eigenvector centrality score, the top seven emerging cited subject categories in 2010 belong to the biomedical sciences meta-discipline.

Discussion and Conclusions

As time progresses and as scientific fields evolve, expand, emerge, and contract, there is a need to maintain and update the search mechanisms and keyword combinations or classifications underlying a search strategy in a particular scientific and technological domain (Thomas et al, 2010). The second version of our nanotechnology search strategy, completed about five years after our initial search approach, reflects and captures changes that have occurred in the nanotechnology domain over this period. We employ an evolutionary approach to updating, in that we maintain a lexical approach but seek to review and add to key inclusion and exclusion terms. The approach to updating leverages both data-intensive analysis and expert input to iterate through candidate keywords and finalize a domain definition.

Our analysis contrasts the updated second search strategy with our initial approach and also seeks to characterize some important shifts in the domain of nanotechnology publications. In terms of total nanotechnology publications identified, the initial and second search strategies identify comparable publication numbers for each year in our panel dataset. That is, notwithstanding the addition of 35 new keywords and thirteen new journals, the aggregate number of publication records has not increased. But there is change within the broad total. The addition of new records is offset by limiting the breadth of contingency search terms deployed along-side microscopy and spectroscopy keywords. The similarity of aggregated publication numbers does not mean that the effort to update the search strategy was not worthwhile. Rather, we judge that the updated search strategy results in both higher recall and precision, enabling greater confidence to be placed in the next round of analyses based on our nanotechnology search approach. Moreover, our comparison of the two search strategies and the results they produce suggest this important observation: while there has been significant expansion in the scale of nanotechnology publication output over the past five or so years, particularly in China but also in other leading developed countries, there has not been a major enlargement in fundamentally new scientific topics as denoted by terms. This is not to say there has been no topic growth: for example, although there was groundbreaking work on graphene prior to 2005, the great expansion of output on this topic has occurred more recently. However, it does seem that the great growth in nanotechnology research since 2005 has occurred mostly within terms and topics that had previously been defined.

Our preliminary review of results obtained from the second version of the nanotechnology search also offers other insights. For example, Roco (2011) has confirmed his model of nanotechnology development as comprising four overlapping generations of research and application: passive nanostructures, active nanostructures, systems of nanosystems, and molecular nanosystems. While the timing of these stages has lagged Roco's earlier predictions, there is some broad evidence that factors underlying nanotechnology generation shifts may be in play. In particular, the development of active nanostructures is conceived as being driven, at least in part, by interest in targeted drugs, biodevices, and other health-related applications. Using the second version of the nanotechnology search approach, our cited subject category analysis shows a pronounced increase in the number of citations to nanotechnology articles in the biomedical sciences, indicating that some shift in knowledge base underlying the corpus of recent nanotechnology research. This corroborates other work (Subramanian et al., 2010) which has used a different bibliometric approach (identifying "active" components) to assess whether there is a shift to active nanostructures.

While we report some early results here, there is plenty of scope for future work both in terms of methodological improvements usable for maintaining and updating bibliometric search strategies and in terms of probing developments in the nanotechnology domain itself. First of all, there is ample opportunity to delve deeper into methodological studies comparing the use of keyword and citation-based analysis as a means to identify a corpus of literature embodied in electronic records. Zitt et al. (2011), for instance, posit that keywords act as overt signals of scientific inquiry whereas citations are more effective in identifying communities of researchers and research streams. However, as De Bellis (2009) observes, although citation analysis is a prominent feature in the study of scientific knowledge output, referencing behavior may be attributed to several causes outside of intellectual critique or hypothesis development. Citations, for example, can refer to methodological insights or even lack substantive merit given the context of mention. A search strategy taking into account these nuances in a field as diverse as nanotechnology may contribute to a more robust dataset with higher recall and precision. At the same time, the benefits of additional complexity must be weighed against portability and the replicability of the search strategy to other data sources (including patents).

A second avenue for advancement in bibliometric analysis, including but not limited to nanotechnology, is in the realm of informatics. Using classification schemes and ontologies, a field's research streams can be described and explored in non-obvious ways. For instance, in bioinformatics, large datasets are organized and categorized in such a way as to introduce the possibility of novel investigation, producing "rescue strategies" whereby failed medical research can be harnessed in more promising future endeavors (Thomas et al, 2010). In nanotechnology, extant research is available en masse in various online indices, but with a more focused data source and concomitant data analysis tools, science and technology scholars would be empowered to perform a number of value-added analyses. Analogous to the rescue strategy in bioinformatics, researchers could, for example, forecast development paths of new and emerging sciences and technologies based on the patterns weaved by existing scholarly work.

A notable consequence of amassing and examining data on scientific output is the production of “metaknowledge”, as defined by Evans and Foster (2011). Metaknowledge allows social scientists to identify models of and antecedents to knowledge production, which is a process shaped by formal and informal channels. We anticipate that our updated and refined nanotechnology search approach will offer a renewed foundational platform from which to study nanotechnologies and the impacts and implications of the ongoing development of this scientific and technological domain and also offer insights for search strategies in other emerging technologies.

Acknowledgements: This research is supported by the Center for Nanotechnology in Society at Arizona State University (National Science Foundation Awards 0531194 and 0937591). The findings in this working paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555-564. doi:10.1016/j.socnet.2007.04.002

Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*, *45*(1), 12-19. doi:10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L

Cunningham, S. & Porter, A. (2011). PICMET.

De Bellis, N. (2009). *Bibliometrics and Citation Analysis*. Lanham, Maryland: Scarecrow Press.

Evans, J. A., & Foster, J. G. (2011). Metaknowledge. *Science*, *331*(6018), 721-725. doi: 10.1126/science.1201765.

Grieneisen, M. (2010). The proliferation of nanotechnology journals. *Nature Nanotechnology*. 7, 273–274.

Grieneisen, M. L., & Zhang, M. (2011). Nanoscience and Nanotechnology: Evolving Definitions and Growing Footprint on the Scientific Landscape. *Small*. doi:10.1002/smll.201100387

Guo, Y., Huang, L., & Porter, A. L. (2010). The research profiling method applied to nano-enhanced, thin-film solar cells. *R&D Management*, *40*(2), 195-208. doi:10.1111/j.1467-9310.2010.00600.x

Ho, S. S., Scheufele, D. A., & Corley, E. A. (2010). Value Predispositions, Mass Media, and Attitudes Toward Nanotechnology: The Interplay of Public and Experts. *Science Communication*, *33*(2), 167-200. doi:10.1177/1075547010380386

Huang, C., Notten, A., & Rasters, N. (2010). Nanoscience and technology publications and patents: a review of social science studies and search strategies. *The Journal of Technology Transfer*, *36*(2), 145-172.

Lakatos, I. (1978). *Philosophical Papers.* Cambridge: Cambridge University Press.

Leydesdorff, L., & Zhou, P. (2007). Nanotechnology as a field of science: Its delineation in terms of journals and patents. *Scientometrics*, *70*(3), 693-713. doi:10.1007/s11192-007-0308-0

Leydesdorff, L., Carley, S., & Rafols, I. (2012). Global maps of science based on the new Web-of-Science categories. Preprint available at: http://arxiv.org/abs/1202.1914

López Cadenas, M. S., Hasmy, A., & Vessuri, H. (2011). Nanoscience and nanotechnology in Venezuela. *Journal of Nanoparticle Research*, *13*(8), 3101-3106. doi:10.1007/s11051-011-0434-8

Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy*, *36*(6), 893-903. doi:10.1016/j.respol.2007.02.005

Newman, M. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5). doi:10.1103/PhysRevE.70.056131

Subcommittee on Nanoscale Science, Engineering and Technology, National Science and Technology Council, Executive Office of the President. (2007). The national nanotechnology initiative: Research and development leading to a revolution in technology and industry.

Pei, R., & Porter, A. L. (2011). Profiling leading scientists in nanobiomedical science: interdisciplinarity and potential leading indicators of research directions. *R&D Management*, *41*(3), 288-306. doi:10.1111/j.1467-9310.2011.00643.x

Porter AL, & Youtie, J. (2009). How interdisciplinary is nanotechnology? Journal of Nanoparticle Research, Vol. 11, No. 5, July 2009, pp. 1023-1041.

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, *10*(5), 715-728. Springer Netherlands. doi: 10.1007/s11051-007-9266-y.

PCAST (2010). Report to the President and Congress on the Third Assessment of the National Nanotechnology Initiative. Executive Office of the President, President's Council of Advisors on Science and Technology. Washington, DC.

PCAST (2012). Report to the President and Congress on the Fourth Assessment of the National Nanotechnology Initiative. Executive Office of the President, President's Council of Advisors on Science and Technology. Washington, DC.

Rafols, I., Porter, A.L., and Leydesdorff, L. (2010) Science overlay maps: A new tool for research policy and library management, *Journal of the American Society for Information Science & Technology*, 61 (9), 1871-1887, 2010

Roco, M. C. (2004). Nanoscale Science and Engineering: Unifying and Transforming Tools. *AIChE Journal*, *50*(5), 890-897. doi:10.1002/aic.10087

Ruhnau, B. (2000). Eigenvector-centrality – a node-centrality? *Social Networks*, *22*(4), 357-365. doi:10.1016/S0378-8733(00)00031-9

Shapira P. and Youtie, J, 2008. Nanodistricts in the United States, *Economic Development Quarterly*, 22(3): 187-199.

Shapira, P., & Wang, J. (2010). Follow the money. *Nature*, 468(7324), 627-8. doi:10.1038/468627a

Shapira, P., J. Youtie, and A.L. Porter. 2010. The Emergence of Social Science Research in Nanotechnology. Scientometrics. 85(2): 595-611.

Subramanian, V., Youtie, J., Porter, A. L., & Shapira, P. (2010). Is there a shift to "active nanostructures"? *Journal of Nanoparticle Research*, *12*(1), 1-10. doi:10.1007/s11051-009-9729-4

Thomas, D. G., Pappu, R. V., & Baker, N. A. (2010). NanoParticle Ontology for cancer nanotechnology research. *Journal of Biomedical Informatics*, *44*(1), 74-59. doi: 10.1016/j.jbi.2010.03.001.

Youtie, J., Porter, AL., Shapira, P., Tang, L., Benn, T. (2011). The Use of Environmental Health and Safety Research in Nanotechnology Research, Journal of Nanoscience and Nanotechnology, 11, 158-166.

Youtie, J., Shapira, P., and Porter, AL., (2008). Nanotechnology publications and citations by leading countries and blocs, Journal of Nanoparticle Research, 10, 981-986.

Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, *42*(6), 1513-1531. doi:10.1016/j.ipm.2006.03.016

Zitt, M., Lelu, A., & Bassecoulard, E. (2011). Hybrid citation-word representations in science mapping: Portolan charts of research fields? *Journal of the American Society for Information Science and Technology*, *62*(1), 19-39. doi:10.1002/asi.21440

Table 1. Nanotechnology definition: Modular search query

Search	**Contingency**	**Terms**
1. Nano*	No	TS=(nano*)
2. Quantum	No	TS=((“quantum dot*” OR “quantum well*” OR “quantum wire*”) NOT nano*)
3. Self-assembly	Yes, MolEnv-I	TS=((“self assembl*” OR “self organiz*” OR “directed assembl*”) AND MolEnv-I)
4. Nano-related	No	TS=(("molecul* motor*" OR "molecul* ruler*" OR "molecul* wir*" OR "molecul* devic*" OR "molecular engineering" OR "molecular electronic*" OR "single molecul*" OR fullerene* OR **buckyball** OR **buckminsterfullerene** OR **C60** OR "**C-60**" **methanofullerene** OR **metallofullerene** OR **SWCNT** OR **MWCNT** OR "coulomb blockad*" OR bionano* OR "langmuir-blodgett" OR Coulombstaircase* OR "PDMS stamp*" OR **graphene** OR "**dye-sensitized solar cell**" OR **DSSC** OR **ferrofluid*** OR "**core-shell**") NOT nano*)
5. Microscopy and spectroscopy	Yes, MolEnv-R	TS=((((TEM or STM or EDX or AFM or HRTEM or SEM or EELS or **SERS** or **MFM**) OR "atom* force microscop*" OR "tunnel* microscop*" OR "scanning probe microscop*" OR "transmission electron microscop*" OR "scanning electron microscop*" OR "energy dispersive X-ray" OR "xray photoelectron*" OR "**x-ray photoelectron**" OR "electron energy loss spectroscop*" OR "**enhanced raman-scattering**" OR "**surface enhanced raman scattering**" OR "**single molecule microscopy**" OR "**focused ion beam**" OR "**ellipsometry**" OR "**magnetic force microscopy**") AND MolEnv-R) NOT nano*)
6. Nano-pertinent	Yes, MolEnv-I	TS=(((NEMS OR Quasicrystal* OR “quasi-crystal*” OR “**quantum size effect**” OR “**quantum device**”) AND MoleEnv-I) NOT nano*)
7. Nano-pertinent	Yes, MolEnv-R	TS=(((biosensor* OR NEMS OR (“sol gel*” OR solgel*) OR dendrimer* OR **CNT** OR “soft lithograph*” OR “**electron beam lithography**” OR “**e-beam lithography**” OR “molecular simul*” OR “**molecular machin***” OR “**molecular imprinting**” OR “quantum effect*” OR “**surface energy**” OR “molecular sieve*” OR “mesoporous material*” OR “**mesoporous silica**” OR “**porous silicon**” OR “**zeta potential**” OR “**epitax***”) AND MolEnv-R) NOT nano*)
8. Nano journals	No	SO=((Fullerene* OR IEEE Transactions on Nano* OR Journal of Nano* OR Nano* OR Materials Science Engineering C* OR **ACS Nano** OR **Current Nanoscience** OR **Digest Journal of Nanomaterials and Biostructures** OR **IEE Proceedings Nanobiotechnology** OR **IET Nanobiotechnology** OR **International Journal of Nanomedicine** OR **International Journal of Nanotechnology** OR **Journal of Biomedical Nanotechnology** OR **Journal of Computational and Theoretical Nanoscience** OR **Journal of Experimental Nanoscience** OR **Nature Nanotechnology** OR **Photonics and Nanostructures*** OR **Wiley Interdisciplinary Reviews Nano***) NOT nano*)
Total		1 or 2 or 3 or 4 or 5 or 6 or 7 or 8

Note: Additions to the initial search strategy (see Porter et al., 2008) are indicated in underlined bold text.

Table 2. Nanotechnology definition: Contingency terms

Contingency	**Terms**
1. MolEnv-I (molecular environment *inclusive*)	(monolayer* or (mono-layer*) or film* or quantum* or multilayer* or (multi-layer*) or array* or molecul* or polymer* or (co-polymer*) or copolymer* or mater* or biolog* or supramolecul*)
2. MolEnv-R (molecular environment *restricted*)	(monolayer* or (mono-layer*) or film* or quantum* or multilayer* or (multi-layer*) or array*)

Table 3. Nanotechnology definition: Exclusion terms

Records containing these terms are removed from the nano* dataset	**Exclude any nano* records containing only one of these terms and no other nano terms**
plankton*	nanometer*
n*plankton	**nano-metre**
m*plankton	nano-meter
b*plankton	**nano-metre**
p*plankton	nanosecond*
z*plankton	nano-second
nanoflagel*	nanomolar*
nanoalga*	nano-molar
nanoprotist*	**nanomole(s)**
nanofauna*	nanogram*
nano*aryote*	nano-gram
nanoheterotroph*	nanoliter*
nanophtalm*	**nanolitre***
nanomeli*	nano-liter
nanophyto*	**nano-litre***
nanobacteri*	
⁂ ~ 270 organism names beginning with nano*	
nano2, nano3, nanos, nanog, nanor, nanoa, nano-, nanog-, nanoa-, nanor-	
nanosatellite*	

Note: Additions to the initial exclusion terms (see Porte et al., 2008) are indicated in underlined bold text. Exclusion terms do not assume wild cards unless * is explicitly indicated.

⁂ See "Supporting Information" for Grinheisen and Zhang, 2011.

Table 4: Comparison of initial and second nanotechnology search strategies (with overlap comparison).

Year	**1990**	**1995**	**2000**	**2005**	**2010**
Second Version (2011-2012) of Nanotechnology Search Strategy					
Publication records	2,383	14,636	25,512	54,329	93,262
Without exclusions	2,646	16,112	27,451	57,059	96,462
Percent change due to exclusions	-11%	-10%	-8%	-5%	-3%
Percent of records matched by nano* (after exclusion terms applied)	9%	22%	39%	65%	76%
Initial Version (2005-2006) of Nanotechnology Search Strategy					
Number of records	2,091	15,757	25,299	55,206	94,257
Comparison between Initial and Second Strategies					
Number of overlapping records with updated search (exclusion terms applied)	1,859	11,622	22,178	50,478	87,778
Percent of overlapping records with updated search (exclusion terms applied)	78%	79%	87%	93%	94%

Source: Analysis of Web of Science publication records.

Note: In 1991, Web of Science changed its journal coverage notably, resulting in a sharp increase in nano-related records. We compared the 1990 with 1991 results and results are comparable on percent of overlapping records with new search (slightly lower at 70%) and percent of nano* records (same 9%), even though the number of records is considerably higher (7,139).

Table 5. Top subject categories in 2010 with corresponding ranks for 2000 and 2005

Subject Category	Records			Rank		
	2000	2005	2010	2000	2005	2010
Materials Science, Multidisciplinary	4,775	11,072	27,385	3	1	1
Physics, Applied	5,648	10,274	19,134	1	2	2
Chemistry, Physical	3,467	7,726	18,655	4	3	3
Chemistry, Multidisciplinary	2,030	6,613	14,888	5	5	4
Nanoscience & Nanotechnology	118	546	14,685	27	25	5
Physics, Condensed Matter	4,992	7,312	13,245	2	4	6
Polymer Science	995	3,070	5,674	9	6	7
Electrochemistry	577	1,454	4,086	14	14	8
Chemistry, Analytical	623	1,519	3,542	13	12	9
Optics	803	1,594	3,334	11	11	10
Physics, Multidisciplinary	1,241	2,216	3,192	7	8	11
Materials Science, Coatings & Films	1,056	1,619	3,033	8	10	12
Engineering, Electrical & Electronic	1,294	2,260	2,972	6	7	13
Engineering, Chemical	414	1,218	2,913	17	15	14
Metallurgy & Metallurgical Engineering	506	1,495	2,883	16	13	15
Physics, Atomic, Molecular & Chemical	943	1,852	2,822	10	9	16
Biochemistry & Molecular Biology	401	1,139	1,954	18	16	17
Materials Science, Biomaterials	94	403	1,753	30	27	18
Chemistry, Inorganic & Nuclear	341	928	1,705	20	19	19
Pharmacology & Pharmacy	202	554	1,682	23	24	20

Source: Analysis of Web of Science publication records using second version of nanotechnology search strategy (see text and Tables 1-3). Exclusion terms applied.

Table 6. Top emerging cited subject categories in 2005 and 2010 with weighted eigenvector scores

Cited Subject Category	Meta-Discipline	*Citations* Records	Rank	*Weighted Eigenvector* Score	Rank
2005					
Engineering, Chemical	Physical Sciences	7,299	1	1.00	1
Environmental Sciences	Environmental Sciences	3,360	4	0.91	2
Engineering, Environmental	Environmental Sciences	2,513	7	0.81	3
Biotechnology & Applied Microbiology	Biomedical Sciences	4,767	2	0.71	4
Materials Science, Biomaterials	Biomedical Sciences	2,919	6	0.49	5
Energy & Fuels	Physical Sciences	2,941	5	0.41	6
Medical Laboratory Technology	Biomedical Sciences	1,519	10	0.38	7
Chemistry, Medicinal	Biomedical Sciences	1,615	9	0.34	8
Plant Sciences	Environmental Sciences	1,309	13	0.34	9
Water Resources	Environmental Sciences	871	16	0.33	10
2010					
Biotechnology & Applied Microbiology	Biomedical Sciences	15,118	2	1.00	1
Pharmacology & Pharmacy	Biomedical Sciences	11,648	4	0.99	2
Engineering, Biomedical	Biomedical Sciences	11,322	5	0.96	3
Materials Science, Biomaterials	Biomedical Sciences	10,787	7	0.94	4
Medicine, Research & Experimental	Biomedical Sciences	5,516	9	0.60	5
Oncology	Biomedical Sciences	5,359	10	0.59	6
Chemistry, Medicinal	Biomedical Sciences	5,786	8	0.56	7
Environmental Sciences	Environmental Sciences	12,047	3	0.44	8
Plant Sciences	Environmental Sciences	4,204	12	0.44	9
Engineering, Multidisciplinary	Physical Sciences	16,171	1	0.41	10

Source: Analysis of Web of Science publication records using second version of nanotechnology search strategy (see text and Tables 1-3). Exclusion terms applied.

Figure 1. Process diagram for working with high-frequency and other keywords

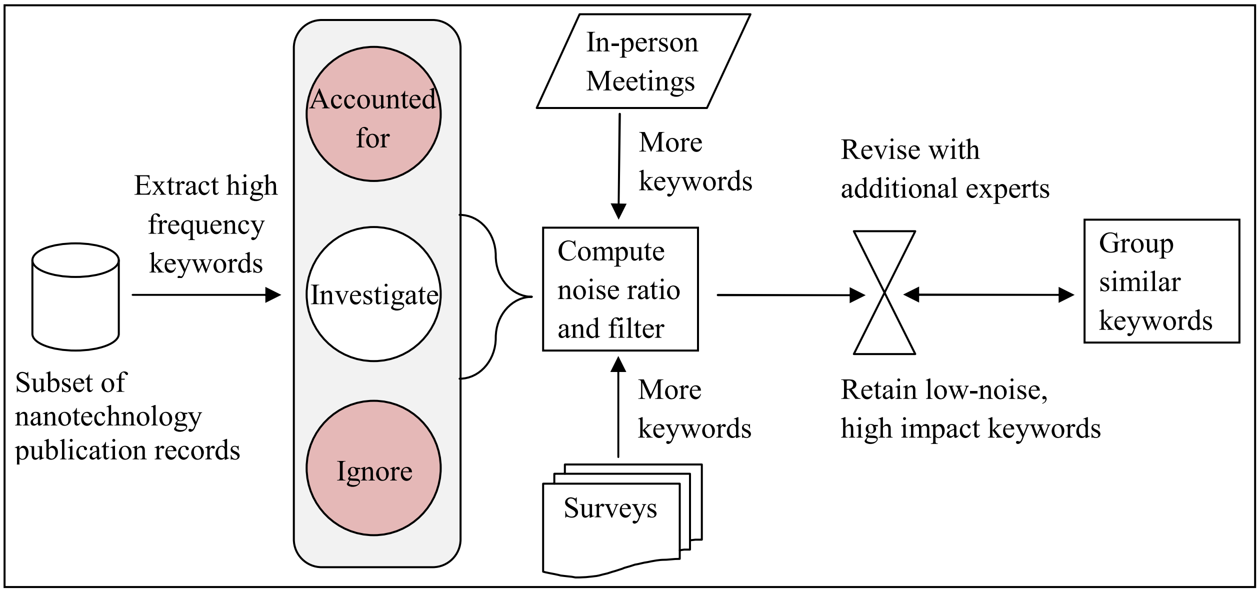
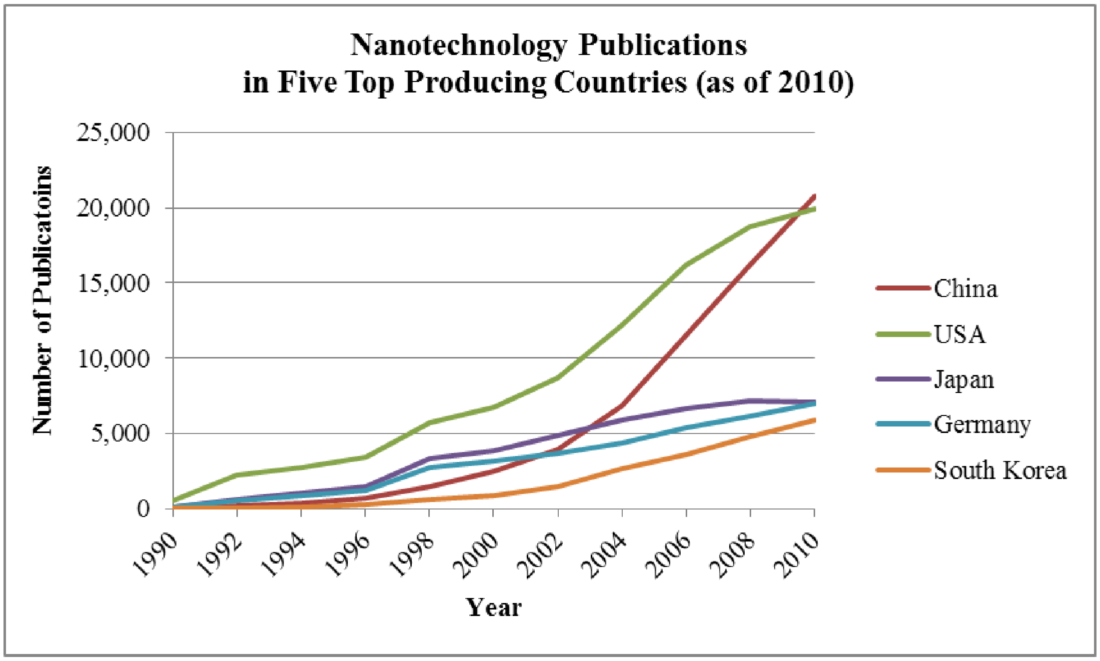
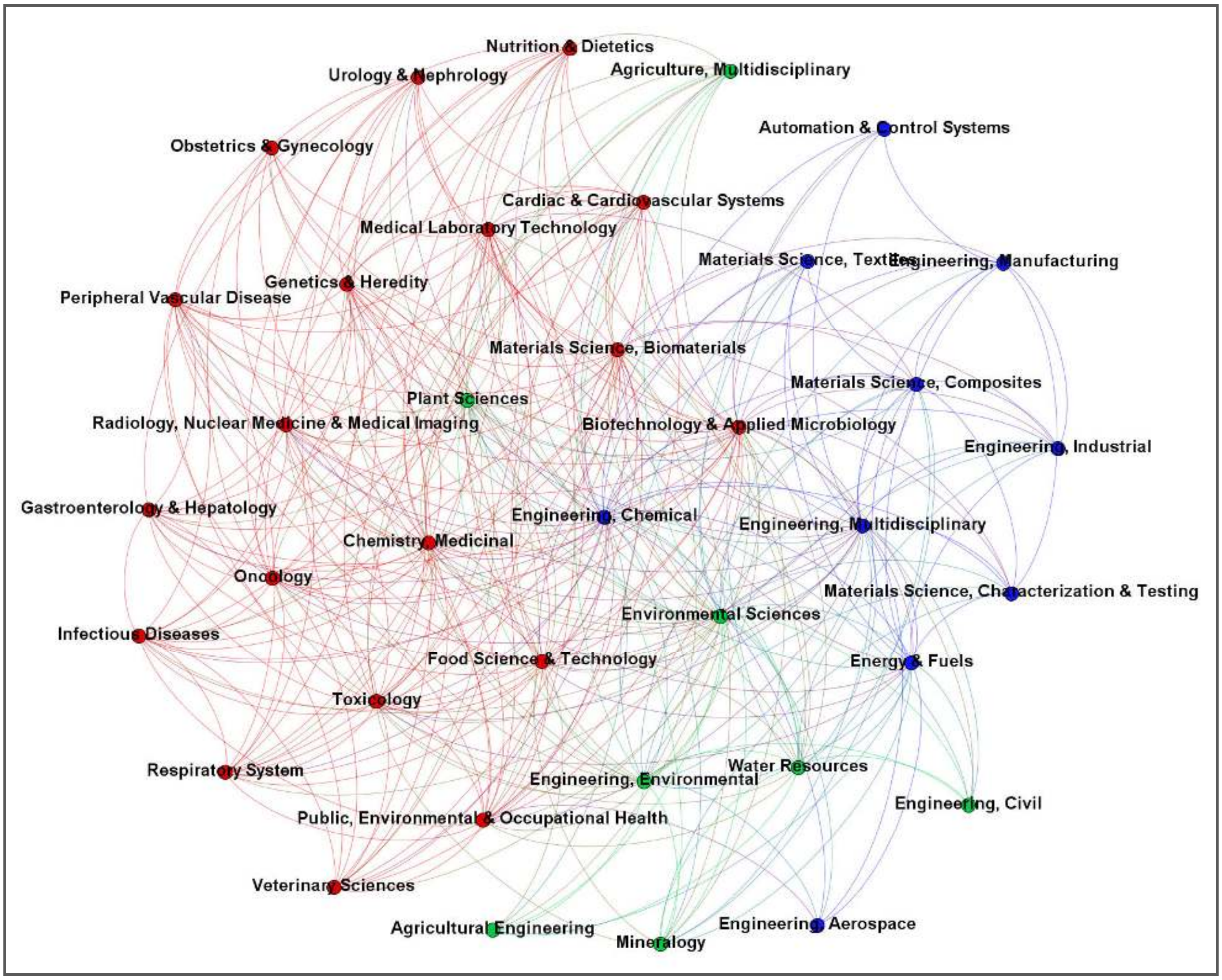


Figure 2. Nanotechnology publications by top producing countries, 1990-2010



Source: Analysis of Web of Science publication records using second version of nanotechnology search strategy (see text and Tables 1-3). Exclusion terms applied.

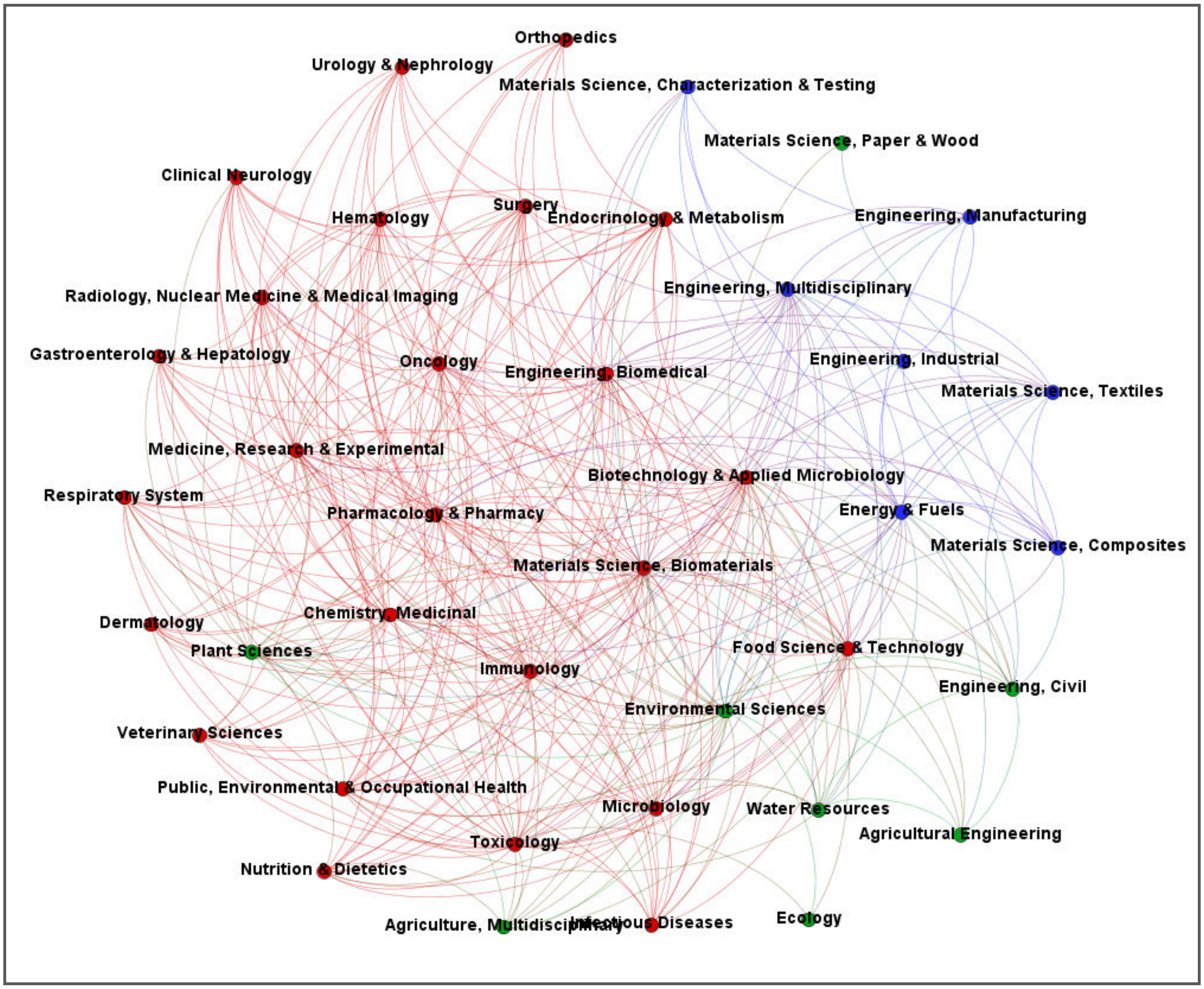
Figure 3. Emerging cited subject categories in 2005



Source: *Source*: Analysis of Web of Science publication records using second version of nanotechnology search strategy (see text and Tables 1-3). Exclusion terms applied.

Based on differences in cited subject category rankings between 2000 and 2005. Shading indicates meta-disciplines: Biomedical Sciences (red), Environmental Sciences (green), and Physical Sciences (blue). Visualized in Gephi using the Fruchterman Reingold layout.

Figure 4. Emerging cited subject categories in 2010



Source: *Source*: Analysis of Web of Science publication records using second version of nanotechnology search strategy (see text and Tables 1-3). Exclusion terms applied.

Based on differences in cited subject category rankings between 2005 and 2010. Shading indicates meta-disciplines: Biomedical Sciences (red), Environmental Sciences (green), and Physical Sciences (blue). Visualized in Gephi using the Fruchterman Reingold layout.

Appendix

	Without contingencies			*With contingencies (inclusive molecular environment*)*		
Search term	**GND**	**Random**	**Ratio**	**GND**	**Random**	**Ratio**
protein corona	3	0				
endocytosis	234	41	40.8%	30	1	*7.8%*
transfection	384	97	58.9%	35	6	40.0%
liposomes	567	47	19.3%	72	7	22.7%
excipent	0	0				
membrane	5,430	997	42.8%	1,290	62	11.2%
core-shell	1,233	27	*5.1%*	360	10	6.5%
ultrafine	561	27	11.2%	69	3	10.1%
sol	35,671	6,511	42.5%	13,351	822	14.4%
molecular crystal	81	6	17.3%	20	3	35.0%

Table 1A: Sample of candidate keywords with noise ratio analysis. Expert solicitation via the survey instrument identified the terms. Two terms shown here were candidates for inclusion in the final search definition: endocytosis (with contingencies) and core-shell (without contingencies). In the end, *endocytosis* would only identify about 30 records, and so it did not make the final cut; *core-shell*, on other hand, appears to be a value added term with a low noise ratio. GND = Georgia Tech nanotechnology dataset. *See Porter et al., 2008.