

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

Sanjay K. Arora · Alan L. Porter · Jan Youtie · Philip Shapira

Received: 17 May 2012 / Published online: 21 December 2012
© Akadémiai Kiadó, Budapest, Hungary 2012

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 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 reviewing and updating this 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 20 years of publication data and analyze the results. Our findings indicate that the updated search approach offers an incremental improvement over the original strategy in terms of recall and precision. Additionally, the updated strategy reveals the importance for nanotechnology of 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.

Keywords Nanotechnology · Bibliometrics · Publications · Search strategy · Cited subject categories

S. K. Arora · A. L. Porter · P. Shapira
School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA

J. Youtie
Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA 30332-0640, USA

P. Shapira (✉)
Manchester Institute of Innovation Research, Manchester Business School, University of Manchester,
Manchester M13 9PL, UK
e-mail: pshapira@mbs.ac.uk

Mathematics Subject Classification 91**JEL Classification** C89 · O30**Introduction**

Nanotechnology involves the understanding and engineering of matter at the nanoscale dimensional range of 1–100 nm. Novel physical, chemical, and biological features can result from the manipulation of nanoscale particles, materials and systems (PCAST 2010). Research in nanotechnology spans a wide spectrum of scientific and technological disciplines including physics, chemistry, material science, engineering and biotechnology.

The inherent characteristics of nanotechnology research and development present challenges for the creation of bibliometric definitions of the field. Size criteria alone are insufficient to distinguish literature in the field (NSTC 2007). Subject category classifications are also inadequate, as nanotechnology diffuses within and across multiple disciplines. Furthermore, journals with nanotechnology (or “nano”) in the publication name do not capture the breadth of the field and may not exclusively focus on nanotechnology (Grieneisen 2010). More sophisticated and nuanced approaches are essential for understanding the evolution of the nanotechnology 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. A key tool has been the development of an encompassing bibliometric 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 (WoS) Science Citation Index (SCI) papers and over 53,000 MicroPatent and INPADOC patent records published between 1990 and mid-2006 (for full details, 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. By mid-2011, our nanotechnology search approach was identifying more than 820,000 WoS 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. These include studies that have identified trajectories of nanotechnology publications and patents (Youtie et al. 2008), nanotechnology research funding sponsorship (Shapira and Wang 2010), active nanotechnology research (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, 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 how well 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 10 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 refresh a bibliometric search strategy and what is gained from updating? In the sizable domain of nanotechnology this is a critical issue, as updating requires significant time and other resource investments. We anticipate that the experience of updating our nanotechnology search approach will be useful in providing insights on this underlying question.

The paper is organized in the following manner. First, we review the original nanotechnology search and associated literature in the context of approaches put forward by other researchers for delineating the nanotechnology domain. The ensuing section presents our updated methodology for identifying research outputs in nanotechnology. We then test the performance of the updated search strategy and explore what it tells us when revisiting nanotechnology publication trends over the last two decades. Finally, we conclude with a discussion of implications and future areas of application.

Context and literature review

Nanotechnology is a science-driven domain that is highly complex and cross-cutting. Researchers have commonly used bibliometrics to monitor trends in the domain beginning with the early use of the term “nano*”, followed by a set of more complex strategies. Recent bibliometric approaches to the nanotechnology domain have taken multi-stage keyword or multi-article citation-based approaches, mirroring the cross-cutting, complex nature of the field itself (Huang et al. 2010).

An important consideration in bibliometric search strategies is 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 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 previous nanotechnology search approach sought to optimize between the extremes of high recall and high precision. We captured a broad array of the nanotechnology literature (thus maximizing recall) while 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.

The initial search strategy consists of two steps (Fig. 1). 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. An eighth modular component searched selected publication sources to capture 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 nano* query where such exclusions will improve precision. We identified 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.

Other researchers have also developed nanotechnology search strategies. 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 core journal strategies. A lexical query relies on expert advice for keyword identification. Although relatively straightforward to implement, the reliability of

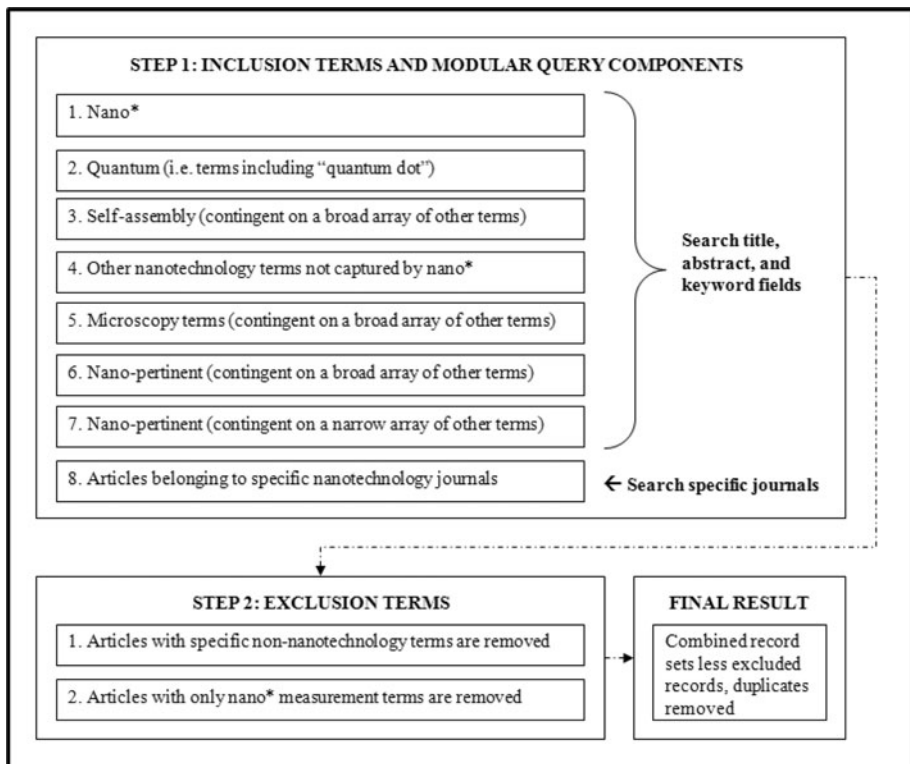


Fig. 1 Overview of nanotechnology search approach. Note: As used by Porter et al. (2008) and in the updated search strategy discussed in the present paper

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.

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. 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. Researcher bias and the influence of the pre-selected keywords on the experts are potential issues with this approach.

A citation-based search strategy relies on a core set of literature to identify other 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. Beginning with a seed set of nanotechnology literature identified by modular queries, Zitt and Bassecoulard (2006) employ citation networks to expand their corpus of nanotechnology publications. 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 definition. Computation and licensing costs are thus salient when considering whether citation analysis is a feasible alternative to lexical (keyword) querying.

Core journal strategies proceed by identifying a nucleus of publications in a scientific field. 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), expand that core set to ten journals. A journal is a “core” publication if it contains “nano” in its title. In theory, precision should be relatively high with this method, but Huang et al. (2010) observe that 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.

None of the aforementioned search strategies for characterizing nanotechnology have previously published subsequent modifications to take into consideration the changing nature of the field. This paper addresses that gap by updating our initial search strategy and reviewing the results. In so doing, we not only present insights about the development of nanotechnology but also offer a methodology that has relevance more broadly in bibliometric strategies to account for change over time in fields of scientific inquiry.

Methodology

As indicated above, the Porter et al. (2008) search strategy produced in 2005–2006 is a lexical approach drawing on expert opinion for keyword identification. Our second or updated version of the search strategy, developed in the period 2011 to early 2012, can be characterized as an evolutionary lexical query, since it employs feedback channels between the keyword identification process and elicitation of expert opinion. Our approach to modifying the list of keywords was informed through (1) systematic, semi-automated evaluations of high-occurrence keywords and (2) interviews, surveys, and other data sources. We discuss each of these techniques below and present the final search specification.

Keyword occurrence analysis

High-frequency keywords that appeared in legacy datasets but were not included in the original Porter et al. (2008) search definition were candidates for the revised search definition. The semi-automated method for identifying such terms requires a measure for comparing candidate terms in the legacy datasets to their respective rate of occurrence in WoS SCI. To produce this measure, we compared the occurrence of high-frequency terms in the abstract, title, and keyword fields of journal articles matching nano* in the benchmark year of 2009 (93,233 records) with their occurrence in a random sample of 40,000 SCI article abstracts, titles, and keywords in the same year. Using VantagePoint—a software application for structured text mining and analysis (see <http://www.theVantagePoint.com>)—we searched for key terms across both datasets. To determine which of the resulting 1,100 candidate terms would warrant additional expert review, we devised a simple noise ratio (η):

$$\eta = (r_y/R_y) / (b_y/B_y)$$

where r = number of hits in random sample, R = total records in random sample, b = number of hits in benchmark nanotechnology dataset, B = total records in benchmark nanotechnology dataset and y = benchmark year.

Eliminating keywords with a noise ratio threshold below 0.200 (or 20 %) produced a list of 75 candidate search terms, some of which could be combined because of obvious lexical similarities or through lemmatization. Most existing terms from Porter et al. (2008) exhibited noise ratios of less than 0.085 (or 8.5 %), so we use this more parsimonious threshold, which yielded ten new keywords that were then moved forward to the next stage of review.

Interviews, surveys, and other data sources

In addition to the semi-automated approach, we identified potential key terms and new journal publication titles using other sources, including nanotechnology press coverage and observations from the Porter et al. (2008) study. We also solicited input from nanotechnology experts through interviews and surveys—with both of these methods being especially important to surfacing and validating candidate query terms.

We began with individual meetings at Georgia Tech with three nanotechnology specialists: a research scientist, a research engineer, and a doctoral candidate. In-depth meetings (typically lasting an hour) were held. One of the interviewees was a manager of a

major nanotechnology user facility and speaker series organizer, giving this individual familiarity with a broad range of nanotechnology research. At this stage, we also piloted a brief expert survey questionnaire. During our conversations and interactions, we gained valuable feedback on new as well as old keywords. We then sent the survey, which asked several questions with respect to the scope and accuracy of our keywords and modular approach, to 67 experts in the US and internationally, including research scientists and academics, industry and government practitioners, and one representative from each of the 14 US National Nanotechnology Infrastructure Network centers. We received 12 completed surveys—a response rate of about 18 %. While low, this response rate is common for voluntary surveys and it was more important to obtain detailed expert review of the search strategy rather than simply a sizable response. To understand whether or not there was response bias, we examined the fields of the participants to ensure that a diversity of fields was represented; this was the case as we received responses from nanotechnology experts in materials, electronics, and biotechnology subfields. The responses either validated or contested some of our additions. For instance, after adding “contact angle,” we received sufficient concern in the survey process to remove the search term in the final updated definition. The in-depth meetings and survey process produced about 100 additional candidate keywords, some of which overlapped with the terms found in the semi-automated 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 0.085 noise threshold requirement.

The updated search specification

The updated search specification maintains the two-step process of the original search. The first step (the inclusion terms) comprises the modular search components as presented in Tables 1 and 2. Additions to the initial definition are emphasized in bold text. The end result from the methods described in the prior sections was the addition of 34 new keywords and 13 new journal titles. 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 WoS). 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, perfluoronanoic, and subnanomolar).

In the second step, the updated version uses a list of exclusion terms to remove unwanted and out-of-scope records (see Table 3). Some exclusion terms, if found in a given record, result in the removal of that record from the dataset *sine qua non*, while 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 “nanosatellite” and spelling variants of measurements at the nanoscale. We also adopted the list of approximately 270 taxonomic organism and species names beginning with nano* (but which are not in themselves nanotechnologies) as identified by Grieneisen and Zhang (2011) (see Table 4).

The final updated search specification, as contained in Tables 1, 2 and 3, was used to search abstract, title, and keywords from WoS SCI records (including journal articles, proceeding papers, news items and reviews) for the inclusive period 1990–2010. The

Table 1 Updated 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 assembly*” OR “self organiz*” OR “directed assembly*”) AND MolEnv-I) |
| 4. Nano-related | No | TS = ((“molecul* motor*” OR “molecul* ruler*” OR “molecul* wir*” OR “molecul* devic*” OR “molecul engineering” OR “molecul electronic*” OR “single molecul*” OR fullerene* OR buckyball OR buckminsterfullerene OR C60 OR “C-60” OR methanofullerene OR metallofullerene OR SWCNT OR MWNT OR “coulomb blockad*” OR bio* 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 Quasicyrystal* 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 “molecul simul*” OR “molecul machin*” OR “molecul imprinting” OR “quantum effect*” OR “surface energy” OR “molecul 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 bold font

Table 2 Updated nanotechnology definition: contingency terms

| Contingency | Terms |
|--|---|
| 1. MolEnv-I (molecular environment <i>inclusive</i>) | (monolayer* OR “mono-layer*” OR film* OR quantum* OR multilayer* OR “multi-layer*” OR array* OR molecu* OR polymer* OR “co-polymer*” OR copolymer* OR mater* OR biolog* OR supramolecul*) |
| 2. MolEnv-R (molecular environment <i>restricted</i>) | (monolayer* OR “mono-layer*” OR film* OR quantum* OR multilayer* OR “multi-layer*” OR array*) |

Table 3 Updated 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* |
| nanophthalm* | 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 Porter et al. 2008) are indicated by bold font. Exclusion terms do not assume wild cards unless * is explicitly indicated

** See Table 4

Table 4 Grieneisen and Zhang's taxonomic exclusion terms

Records containing these terms are removed from the nano* dataset

Nanoa; Nanoacalles; Nanoagraylea; nanoapiculatum; Nanoarchaea; Nanoarchaeota; Nanoarchaeum; Nanoastegotherium; Nanobagrus; Nanobalcis; Nanobaris; Nanobates; Nanobatinae; Nanobius; Nanobryaceae; nanobryoides; Nanobuthus; Nanocalcar; Nanocambridgea; nanocapillare; nanocarpa; nanocarpum; Nanocarpus; Nanocassiope; Nanocavia; nanocephalum; Nanocheiroidon; Nanochilina; Nanochilus; Nanochitina; Nanochlaenius; Nanochlorum; Nanochoerus; Nanochromis; Nanochrysopa; Nanochthonius; Nanocixius; Nanocladius; Nanoclarelia; Nanoclaveia; nanoclimacium; Nanoclymenia; Nanocnide; Nanocochlea; Nanocolletes; Nanocondylodesmus; Nanocopia; Nanocoquimba; Nanocrinus; Nanoctenus; Nanocthispa; Nanocuridia; Nanocurius; Nanocyclopia; Nanocynodon; Nanocythere; Nanodacna; nanodactylus; Nanodamon; Nanodea; nanodealbata; Nanodectes; nanodella; Nanodelphys; nanodendron; nanodes; Nanodiaparsis; Nanodiaptomus; Nanodidelphys; Nanodiella; Nanodiodes; Nanodiplosis; Nanodiscus; nanodisticha; Nanodromia; Nanodynerus; Nanofila; Nanofiliidae; Nanogalatea; nanoglobum; Nanoglossa; Nanognathia; Nanognathus; Nanogomphodon; Nanogona; Nanogonals; Nanogorgon; Nanogramma; Nanograptus; Nanogyra; Nanogyrini; Nanohalus; Nanohammus; Nanohemicerca; nanohystrix; nanoïdes; Nanoini; Nanojapyx; Nanokerala; Nanokermes; Nanola; Nanolachesilla; Nanolania; Nanolauthia; Nanolestes; Nanolichus; Nanolobus; Nanoloricida; Nanolpium; nanolumen; Nanomaja; Nanomantinae; Nanomantini; Nanomantis; Nanomelon; Nanomermis; Nanomerus; Nanomeryx; Nanometa; Nanometidae; Nanometinae; Nanometra; Nanomia; Nanomias; Nanomicrophytes; Nanomilleretta; Nanomimus; Nanomis; nanomitra; Nanomitriella; Nanomitriopsis; Nanomitus; Nanomutilinae; Nanomutilla; Nanomyces; Nanomyina; Nanomyrmacyba; Nanomyrme; Nanomys; Nanomysis; Nanomysmena; Nanonaucoris; Nanonavis; Nanoneis; Nanonemoura; nanonocticolus; Nanonycteris; Nanopachyiulus; Nanopagurus; Nanopareia; Nanoparia; Nanopatula; nanopennatum; Nanoperla; Nanophareus; Nanophemera; Nanophthalmus; Nanophya; Nanophydes; Nanophyidae; Nanophydini; Nanophytes; Nanophytinae; Nanophyetus; Nanophyidae; Nanophyinae; Nanophyini; nanophylla; Nanophyllini; Nanophyllium; nanophyllum; nanophyllus; Nanophytes; nanophyti; Nanophyton; Nanopilumnus; Nanopitar; Nanoplagia; Nanoplax; Nanoplaxes; Nanoplectrus; Nanoplinthisus; Nanopodella; Nanopodellus; nanopolymorphum; Nanopolystoma; Nanopria; Nanops; Nanopsallus; Nanopsis; Nanopsocetae; Nanopsocus; Nanopterodectes; Nanopterum; Nanoptilium; Nanopus; nanopyxis; Nanoqia; nanoqsunquak; Nanorafonus; Nanorana; Nanoraphidia; Nanorchestes; Nanorchestidae; Nanorhamphus; Nanorhathymus; Nanorhopaea; Nanorrhacus; Nanorrhynchus; Nanorthidae; Nanorthis; Nanos; nanosalicium; Nanosauridae; Nanosaurus; Nanoschema; Nanoschetus; Nanoscydmus; Nanoscypha; Nanosella; Nanosellini; nanoserranus; Nanosesarma; nanosetus; Nanosilene; Nanosiren; Nanosius; Nanosmia; Nanosmilus; nanosomus; nanospadix; nanospathulatum; Nanospira; Nanospondylus; nanospora; Nanosteatoda; nanostellata; Nanostictis; Nanostoma; Nanostomus; Nanostrangalia; Nanostrea; Nanostreptus; Nanosura; Nanosylvanella; Nanotagalus; Nanotanaupodus; nanotaphus; Nanotermitodius; Nanothamnus; nanothecioidea; Nanothecium; Nanothinophilus; Nanothrips; Nanothyris; Nanotitan; Nanotitanops; Nanotopsis; Nanotragulus; Nanotragus; Nanotrema; Nanotrepes; Nanotrigona; Nanotriton; Nanotrombium; Nanotyrannus; Nanoviridae; Nanovirus; Nanowana; Nanowestratia; Nanoxolocopa.

Source: “Supporting Information” for Grieneisen and Zhang (2011). Terms starting with lower-case “n” are species; those starting with an upper case “N” are higher taxonomic classifications

analyzed results are helpful not only in assessing the performance of the updated strategy vis-à-vis the initial definition but also to characterize trends and patterns in the field in substantive ways, as discussed in the following section.

Results

We present the results of the updated search in three subsections: a comparison of the performance of the initial and updated 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.

Table 5 Comparison of initial and updated nanotechnology search strategies

| Search strategy | 1990 | 1995 | 2000 | 2005 | 2010 |
|--|-------|--------|--------|--------|--------|
| Initial version (2005–2006) of nanotechnology search strategy (Porter et al. 2008) | | | | | |
| Number of records | 2,091 | 15,757 | 25,299 | 55,206 | 94,257 |
| Updated version (2011–2012) of nanotechnology search strategy | | | | | |
| Publication records without exclusions | 2,646 | 16,112 | 27,451 | 57,059 | 96,462 |
| Publication records with exclusions | 2,383 | 14,636 | 25,512 | 54,329 | 93,262 |
| Percent change after exclusions (%) | −11 | −10 | −8 | −5 | −3 |
| Percent of records matched by nano* (after exclusion terms applied) (%) | 9 | 22 | 39 | 65 | 76 |
| Comparison between initial and updated 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 expanded its journal coverage. Although the number of records (7,139) identified for 1991 by our search strategy is considerably higher than in the previous year, the 1990 and 1991 results are comparable on percent of overlapping records with new search (slightly lower at 70 %) and percent of nano* records (same at 9 %)

Performance

A comparison of the result sets returned from the initial and updated search queries reveals a significant overlap in the number of records identified in any given year (Table 5). 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 updated 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, than publications from the 2000s. In other words, the nano* prefix tends to capture more papers not relevant to nanotechnology in the 1990s than in the 2000s. This conclusion is based on term searching in the abstract, title, and keywords. The full text of the publication is not examined; this caveat indicates a limitation of the finding and a pathway for future research. Within this limitation, the findings suggest that over time a majority of researchers have arrived at broad overall understanding of what is nanotechnology. 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.

The initial and updated 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 5). Our updated search strategy identifies about 760,000 WoS nanotechnology papers published between 1990 and 2010; of these, some

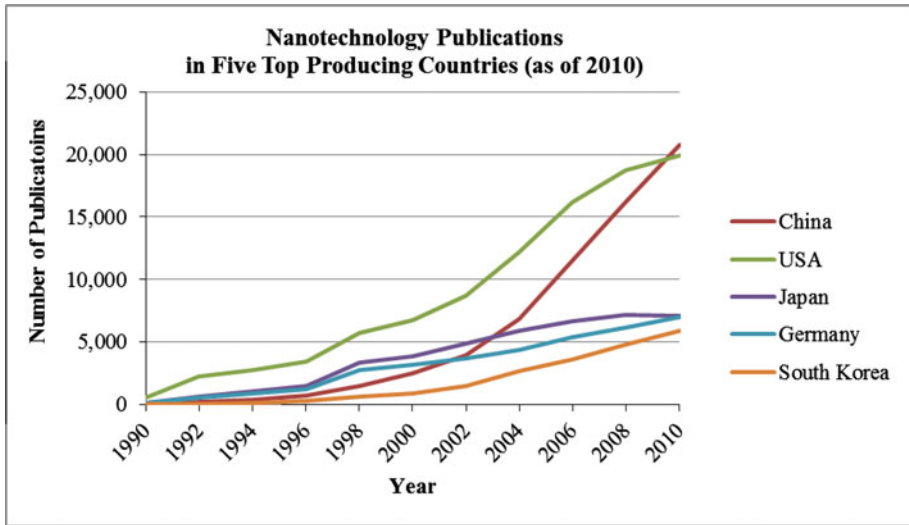


Fig. 2 Nanotechnology publications by top producing countries, 1990–2010. Source: Analysis of Web of Science publication records using updated version of nanotechnology search strategy (see text and Tables 1, 2, 3). Exclusion terms applied

2,400 were published in 1990, with over 93,000 published in 2010. Most years in this time period 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 countries across our 20 year time horizon (see Fig. 2).

The top two nanotechnology research publishing countries are the US and China. Both nations initiated national nanotechnology initiatives at about the same time in the early 2000s (Shapira and Wang 2010), and 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. Yet, publication counts do not necessarily equate to publication influence (Youtie et al. 2008). Articles with authors based in the US and the 27 member countries of the European Union each account for nearly 35 % of citations to the world's nanotechnology articles, while articles with Chinese authors account for only about 20 % of citations. Analysis of our results finds that about 40 % of Chinese authors garner no citations, compared with 29 % of US papers. If papers in leading journals such as *Science* and *Nature* (each with WoS journal impact factors of more than 30), and *Proceedings of the National Academy of Science* (WoS journal impact factor of nearly 10) are considered, the US continues to

maintain a significant leadership in producing high impact nanotechnology papers ahead of all other nations (PCAST 2012).

One caution is that papers from non-English speaking countries are likely to be under-represented in the results. Some 98 % of the identified WoS papers in nanotechnology in 2010 were published in English, although 23 other languages were also represented, with Chinese accounting for 75 % of the non-English publications. This finding is consistent with the work of Lin and Zhang (2007) concerning the rise of Chinese WoS language publications. However, it is not clear how much nanotechnology research in non-English speaking countries is overlooked by relying on WoS. For example, Shapira and Wang (2009) observe that incentives for academic qualifications and for career development in China increasingly direct Chinese researchers to publish in WoS indexed journals.

Emerging research areas

To identify emerging research areas in nanotechnology, we turn to analyzing the subject categories and cited subject categories of publication records as identified by the updated search query. Subject categories are based on classifications of journals used in WoS, drawing on the science mapping method developed by Leydesdorff et al. (2012). We analyze the top 20 subject categories in 2010 and compare how these rankings have changed since 2000 and 2005 (Table 6). Many of the relative rankings remain the same over this 10 year time period. For example, “Materials Science, Multidisciplinary”, “Physics, Applied”, “Chemistry, Physical”, and “Chemistry, Multidisciplinary” are consistently in the top five subject categories. However, two noticeable trends are evident. First, “Nanoscience and Nanotechnology”, introduced into WOS in 2005, not surprisingly reflects rapid growth; as of December 2011, 27 journal titles in the WoS SCI and 66 journals in SCI Expanded belong to this subject category. Second, 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 and Chemical” and “Engineering, Electrical and 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.

These observations should be interpreted with caution. Subject categories are applied at the journal level, and all articles in a publication title inherit these classifications. However, not all articles in a journal align with its assigned subject category. The addition of new journals in a particular subject area also can skew 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 draws 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 and then apply a thesaurus to obtain the corresponding cited subject categories. This approach gives us a proxy for the “research programme” concept initially described by Lakatos (1978). This consists of a hard core of assumptions and a protective belt which shapes and advances “problem shifts” (or movement to new successor theories). It is via changes in the protective belt that we seek to explore nanotechnology’s most recent problem shifts, and we do this by analyzing transitions in cited subject categories.

Table 6 Top nanotechnology subject categories in 2010 with corresponding ranks for 2000 and 2005

| Subject category | Nanotechnology publications | | | 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 and 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 and Films | 1,056 | 1,619 | 3,033 | 8 | 10 | 12 |
| Engineering, Electrical and Electronic | 1,294 | 2,260 | 2,972 | 6 | 7 | 13 |
| Engineering, Chemical | 414 | 1,218 | 2,913 | 17 | 15 | 14 |
| Metallurgy and Metallurgical Engineering | 506 | 1,495 | 2,883 | 16 | 13 | 15 |
| Physics, Atomic, Molecular and Chemical | 943 | 1,852 | 2,822 | 10 | 9 | 16 |
| Biochemistry and Molecular Biology | 401 | 1,139 | 1,954 | 18 | 16 | 17 |
| Materials Science, Biomaterials | 94 | 403 | 1,753 | 30 | 27 | 18 |
| Chemistry, Inorganic and Nuclear | 341 | 928 | 1,705 | 20 | 19 | 19 |
| Pharmacology and Pharmacy | 202 | 554 | 1,682 | 23 | 24 | 20 |

Source: Analysis of Web of Science publication records using updated version of nanotechnology search strategy (see text and Tables 1, 2, 3). Exclusion terms applied

To accomplish this, 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). We compare the rank order of cited subject categories from one period to the next, focusing on those cited subject categories that experience a variation of four or more (positive) positions. 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 exclude “Nanoscience and Nanotechnology” due to its recent inclusion into the WoS typology and because of its role as an all-encompassing cited subject category, and also ignore cited subject categories with fewer than 500 total citations in the three year sample. To visualize the progressivity of the “research programme”, 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 nanotechnology’s emerging cited subject categories in 2005 (see Fig. 3), by meta-discipline, depicts a strong presence of subject categories related to biomedical

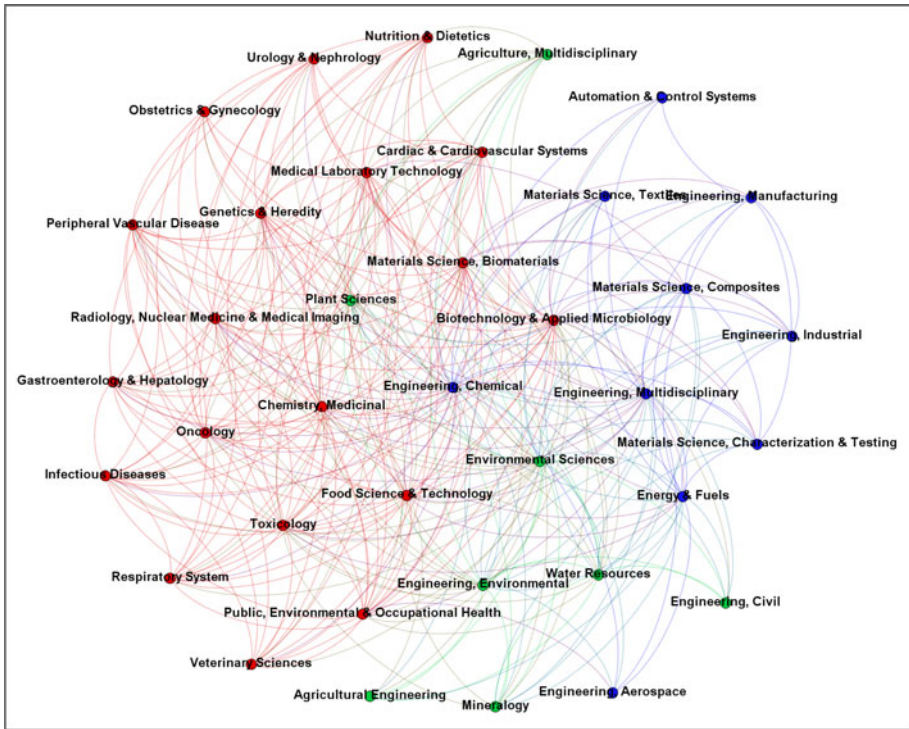


Fig. 3 Emerging nanotechnology cited subject categories in 2005. Source: Analysis of Web of Science publication records using updated version of nanotechnology search strategy (see text and Tables 1, 2, 3). Exclusion terms applied. Note: Based on differences in cited subject category rankings between 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

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 WoS 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 Fig. 4). The physical sciences and environmental sciences each maintain seven and eight emerging cited subject categories, respectively. 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, the analysis highlights potential emerging areas of nanotechnology knowledge in recent years.

The network diagrams provide us with a summary level overview of how different up-and-coming nanotechnology cited subject categories align and connect; however, the visualizations do not confer precise indicators of importance and weight. Consequently, we turn to two other measures: the number of citations to a particular subject category and eigenvector weighted centrality (see Table 7). Whereas number of citations reveals the number of references to articles in the emergent subject category, weighted eigenvector

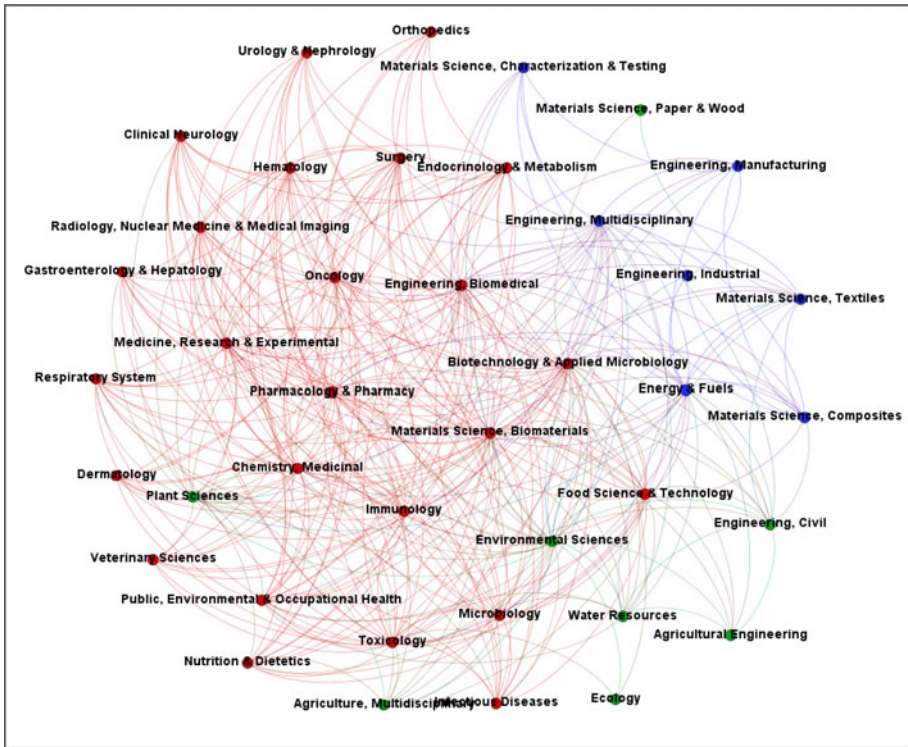


Fig. 4 Emerging nanotechnology cited subject categories in 2010. Source and notes: See Fig. 3

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 (Ruhnau 2000).

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 (Fig. 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 and Applied Microbiology”, and so on. A 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.

Table 7 Top emerging nanotechnology cited subject categories in 2005 and 2010

| Cited subject category | Meta-discipline | Citations | | Weighted Eigenvector | |
|--|------------------------|-----------|------|----------------------|------|
| | | Records | Rank | 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 and Applied Microbiology | Biomedical Sciences | 4,767 | 2 | 0.71 | 4 |
| Materials Science, Biomaterials | Biomedical Sciences | 2,919 | 6 | 0.49 | 5 |
| Energy and 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 and Applied Microbiology | Biomedical Sciences | 15,118 | 2 | 1.00 | 1 |
| Pharmacology and 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 and 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: Same as Table 6

Using the same framework for investigation, the 2010 network, in conjunction with the top ten cited subject categories, implies that progressive problem shifts in nanotechnology 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 nanotechnology cited subject categories, such as “Pharmacology and 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

Scientific fields evolve, expand, emerge, and contract over time. For bibliometric analysis, this implies the need to maintain and update the mechanisms, keyword combinations and classifications underlying search strategies for targeted scientific and technological

domains (Thomas et al. 2010). The updated version of our nanotechnology search strategy, completed about 5 years after our first search approach, seeks to reflect and capture 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 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 updated search strategies identify comparable publication numbers for each year in our panel dataset. That is, notwithstanding the addition of 34 new keywords and 13 new journals, the aggregate number of publication records has not increased. 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 another important observation. 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. However, there has not been a major enlargement in fundamentally new scientific topics not captured by nano*. 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. Additionally, while there may be new concepts emerging, they generally appear to be captured by terms beginning with the “nano” prefix. Nonetheless, it does seem that the great growth in nanotechnology research since 2005 has occurred mostly within terms and topics that had previously been defined.

Further insights are discernible from the updated search results. For example, Roco (2004) proposes a 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 early 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 updated 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 early results here, there remains 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.

One 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 nanotechnology search approach will offer a renewed foundational platform from which to study nanotechnologies. We hope that the updated approach will advance assessment of 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.

Acknowledgments 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.
- Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, 45(1), 12–19.
- Cunningham, S. & Porter, A. (2011). Bibliometric discovery of innovation and commercialization pathways in nanotechnology. Proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET 2011), Portland OR. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6017795.
- De Bellis, N. (2009). *Bibliometrics and citation analysis*. Lanham: Scarecrow Press.
- Evans, J. A., & Foster, J. G. (2011). Metaknowledge. *Science*, 331(6018), 721–725.
- 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*, 7(20), 2836–2839.

- Huang, C., Notten, A., & Rasters, N. (2010). Nanoscience and technology publications and patents: a review of social science studies and search strategies. *Journal of Technology Transfer*, 36(2), 145–172.
- Lakatos, I. (1978). The methodology of scientific research programmes. In J. Worrall, G. Currie (Ed.), *Philosophical papers*. Cambridge: Cambridge University Press.
- 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>.
- Leydesdorff, L., & Zhou, P. (2007). Nanotechnology as a field of science: its delineation in terms of journals and patents. *Scientometrics*, 70(3), 693–713.
- Lin, M., & Zhang, J. (2007). Language trends in nanoscience and technology: the case of Chinese-language publications. *Scientometrics*, 70(3), 555–564.
- 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.
- Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5), 1–9. (056131).
- NSTC (2007). The National Nanotechnology Initiative: research and development leading to a revolution in technology and industry. Subcommittee on Nanoscale Science, Engineering and Technology, National Science and Technology Council. Washington: Executive Office of the President. http://www.sandia.gov/NINE/documents/NNI_Strategic_Plan_2007.pdf
- PCAST (2010). Report to the President and Congress on the Third Assessment of the National Nanotechnology Initiative. President's Council of Advisors on Science and Technology. Washington: Executive Office of the President. <http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports>
- PCAST (2012). Report to the President and Congress on the Fourth Assessment of the National Nanotechnology Initiative. President's Council of Advisors on Science and Technology. Washington: Executive Office of the President. <http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports>
- Porter, A. L., & Youtie, J. (2009). How interdisciplinary is nanotechnology? *Journal of Nanoparticle Research*, 11(5), 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.
- Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: a new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9), 1871–1887.
- Roco, M. C. (2004). Nanoscale science and engineering: unifying and transforming tools. *AIChE Journal*, 50(5), 890–897.
- Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality? *Social Networks*, 22(4), 357–365.
- Shapira, P., & Wang, J. (2009). From lab to market? Strategies and issues in the commercialization of nanotechnology in China. *Journal of Asian Business Management*, 8(4), 461–489.
- Shapira, P., & Wang, J. (2010). Follow the money. *Nature*, 468(7324), 627–628.
- Shapira, P., & Youtie, J. (2008). Nanodistricts in the United States. *Economic Development Quarterly*, 22(3), 187–199.
- 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.
- Thomas, D. G., Pappu, R. V., & Baker, N. A. (2010). NanoParticle ontology for cancer nanotechnology research. *Journal of Biomedical Informatics*, 44(1), 59–74.
- Youtie, J., Shapira, P., & Porter, A. L. (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 and Management*, 42(6), 1513–1531.
- 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.