Working paper, see : Delemarle, A., Kahane, B., Villard, L. & Laredo, P. (2009). *Geography of knowledge production in nanotechnologies: a flat world with many hills and mountains.* Nanotechnology Law & Business, 6(1).

Geography of knowledge production in nanotechnologies : A flat world with many hills and mountains

AURELIE DELEMARLE*, BERNARD KAHANE**, LIONEL VILLARD*** and PHILIPPE LAREDO****

Abstract: What kind of world is the one in which scientific production in nanotechnology takes place? Are localizations and hierarchies related to previous technology waves (ICT, biotechnologies) or do new places emerge? Using a specific 1998-2006 database of more than 538,000 publications (Mogoutov & Kahane, 2007), we take a cluster approach to show that activities concentrate at more than 90% in 200 geographic clusters. This strong concentration at clusters' level reveal an uncommon world picture where national frontiers are challenged while new actors, mainly located in Asia, display strong presence and dynamic. We then analyze how clusters visibility and thematic presence relates to their size and growth. Consequences of this world-clustered picture for public policy are then discussed.

Introduction

Research on nanoscale phenomenom has increased everywhere. Publications in fields related to nanotechnologies have increased by 12% per year between 1998 and 2006. Attracted by the promises of the technology to come programs in nanosciences flourished in almost every country of the world. Nanotechnology being considered as a generic technology of the 21th century, resources invested are considerable, especially public resources: an estimate of $1,780 million in the US in 2007, $975 million in Japan, $ 563 million in German public programs[1], $222 million in South Korean for the same year etc. What kind of continental, national and subnational landscape and hierarchies do these investments shape?

To answer these questions and to analyze the development of nanosciences at the world level, we developed and use automatized methods that are not expert dependent.We first detail our methodology: why and how the database on which this work is based was constructed, followed by how geocoding and geographical clusterization of these data were made. Later, we describe how geographical clustering was performed. These allow us to first show that, while the world may be indeed globalized[2] and thus flat[3], its landscape is nevertheless shaped by a limited number of hills and mountains. Thus, as wine, nanotechnologies do not grow

---

* Aurélie Delemarle is an assistant professor at Université Paris-Est / ESIEE Management and a research fellow at LATTS – IFRIS (Laboratoire Techniques, Territoires et Sociétés – Institut Francilien Recherches Innovations Sociétés)
** Bernard Kahane is a professor at Université Paris-Est / ESIEE Management and a research fellow at LATTS – IFRIS
*** Lionel Villard is a research assistant (???) at Université Paris-Est / ESIEE Management and a research fellow at LATTS – IFRIS
**** Philippe Larédo is a scientific professor at Université Paris-Est / ESIEE Management and a research fellow at LATTS – IFRIS and a professor at Manchester University (Manchester Business School, Manchester Institute of Innovation Research and coordinator of the PRIME European network of Excellence (www.prime-noe.org)

[1] Technology Transfer Centre, 2007

[2] In line with the knowledge-based society promoted (Lisbon, 2000) and the growing use of ICT in the diffusion of knowledge.

[3] Thomas L. Friedman, The World Is Flat: A Brief History of the Twenty-First Century, Farrar, Straus and Giroux, 2005,

everywhere but seem to need both good "soil" (starting *"favorable"* conditions) and dedicated "care" (sustained investment). Second, it gives testimony of the growing importance of new players with top ranked and highly dynamic clusters located in countries (China, South Korea, Taiwan, and Singapore) that are not part of the geographic triad as it is defined by the World Bank[4]. Third, it highlights how clusters differ in their visibility and thematic equilibrium and points notably to the importance of size and thus critical mass. Consequences for public policy are then discussed.

I. DATA EXTRACTION

Describing the development of nanosciences is dependent of the quality of data at hands. Tracking publications in scientific journals is seen as the easiest and most relevant way to work on science production[5]. Nevertheless, emerging fields such as nanotechnologies bring a specific challenge since publications in nanoscience are present in an extended number of journals where they are mixed with non nanoscience papers. Meanwhile, in traditional databases such as the Web of Science (WoS), a tag to mark papers dealing with nanoscience does not yet exist. Using relevant keywords is also not a substitute. Indeed, as scientists engaged in the field try progressively to define their language, relevant keywords emerge and die in a Darwinian process in which only the strongest survive[6].Summing up all these elements, the delineation of nanotechnologies' relevant articles is not an easy task. Indeed, building a set of data, which is both specific and offers extensive coverage of the field has been a intense challenge since nanosciences and nanotechnologies have gained attraction. At first, the delineation of nanoscience related publicatiosn made using what can be described as a nanostring, that is a list of words that include nano (e.g. nanotube) while excluding those with no connection to the field (e.g. nanosecond)[7]. Fraunhofer-ISI later extended on this approach by using experts to provide a list of key words that would complement the initial query. While the first approach was limited, the second was highly expert dependent, meaning its specificity and rationales could not be easily tracked. Further, while nano science was suspected to expand drastically, no one was sure that this query was keeping in line and was still covering most of the nano field. Meanwhile, other sophisticated approaches were experimented[8] trying to get rid of expert bias and to provide a better coverage of the nano

---

[4] World Bank, World development report, 2006. The term triad was first used by Kenichi Ohmae, Triad Power: The Coming Shape of Global Competition, 1985

[5] Michel Callon, John Law and Arie Rip, Mapping the Dynamics of Science and Technology, Macmillan, London, 1986.

[6] Andrea Bonaccorsi, Search regimes and the industrial dynamics of science. Paper presented at the PRIME General Assembly, Manchester. Last accessed, November 2nd, 2008 at http://www.prime-noe.org/index.php?project=prime&locale=en&level1=menu1_prime_1b8057d059a36720_1&level2=6&doc=Annual_Conference&page=3.

[7] William M Tolles,. National Security Aspects of Nanotechnology. In: Roco, M.C. and W.S. Bainbridge (eds.) (2001) *Societal Implications of Nanoscience and Nanotechnology,* NSET Workshop Report. Arlington: National Science Foundation. Available at http://www.wtec.org/loyola/nano/NSET.Societal.Implications/, last accessed on November 2, 2008

Martin Meyer et al. , *Mapping Excellence in Nanotechnologies Preparatory Study*. Report for the Directorate-General Research, 2001,. Luxembourg: European Commission. Available at http://ec.europa.eu/research/era/pdf/nanoexpertgroupreport.pdf, last accessed on November2, 2008.

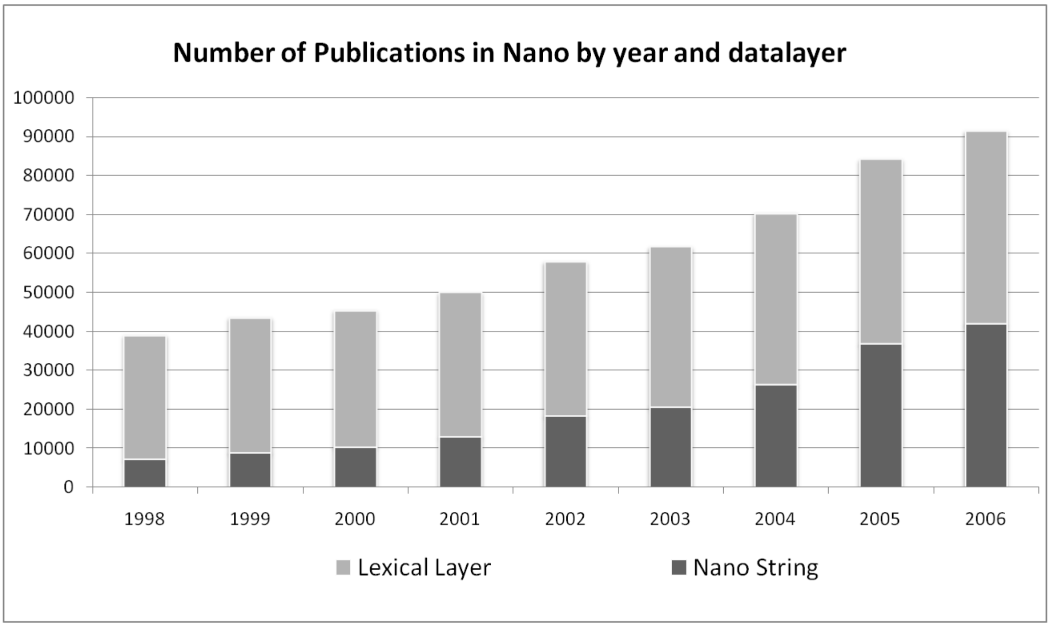
Steve Dunn, S. and Roger W. Whatmore ,*Nanotechnology advances in Europe*. Directorate-General for Research, Scientific and Technological Options Assessment Series. 2002,Working Paper STOA 18 EN. Luxembourg: European *Parliament.*

T Braun, A. Schubert and S. Zsindely, Nanoscience and nanotechnology on the balance. *Scientometrics, 1997,* 38(2): 321-325

[8] Zitt and Bassecoulard provide an analysis using inter-citations.

field. Indeed, they arrived to a probably unique results, however requiring a very specific access to the WoS only a few teams can have in the world. This forbade any capacity for cross-comparisons and shared analyses. Mogoutov and Kahane[9] have been the first to propose an automated lexical modular methodology instead of using experts[10]. Their method is based on an initial nanostring that is progressively enriched by other keywords selected using an inter-citation network density method. New keywords are tested for their specificity and added up to form the final query[11]. This approach allows us to treat a very large number of entries, one of the largest so far in the bibliometrics field[12]. We used this modular query from 1998 to 2006 from the WoS arriving to a total amount of 538 000 publications. This shows the extent of worldwide knowledge production in nanosciences for the period and provides testimony of its growth during the 1998-2006 period. It also shows that the nanostring commonly used to track emergence is incomplete[13] (see Figure 1)

Figure 1 – Comparaison of coverage of the nanoscience field based on different extraction methodology.



Michel Zitt and Elise Bassecoulard, Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. Information Processing and Management, 2006, 42: 1513–1531

[9] Mogoutov A, Kahane B., Data Search Strategy for Science and Technology Emergence : A Scalable and Evolutionary Query for Nanotechnology Tracking, Research Policy 36 (6), 2007

[10] Other studies have been following :

Ronald N. Kostoff, Raymond G. Koytcheff, and Clifford G.Y. Lau, Global nanotechnology research metrics. Scientometrics, 2007, 70(3): 565-601.

Lynne G. Zucker, Michael R. Darby, Jonathan Furner, Robert C. Liu, Hongyan Ma , Minerva unbound: Knowledge stocks, knowledge flows and new knowledge production. Research Policy, 2007, 36: 850–863

Alan L. Porter, Jan Youtie, Philip Shapira and David J. Schoeneck, D. Refining Search Terms for Nanotechnology', Journal of Nanoparticle Research, 2008, vol. 10/5, pp. 715–728

[11] We have used the same method to build a database of patents using Patstat (avril 2008) in order to analyse the development of nanotechnology over the same period of time.

[12] See for a comparison of approaches : See Can Huang, Ad Notten ad Nico Rasters, Nanotechnology Publications and Patents: A Review of Social Science Studies and Search Strategies, 2008, WP#2008-058, UNU-MERIT

[13] The coverage of the field by using only the nanostring is low at the beginning of the period. It progressively increases.

II. DATA GEOCODING

We extracted addresses of each publication and localized them at the level of the city (via Map Point and ArcGIS). For that matter, we developed a program based on the identification of cities, of states for federal countries and of prefectures for Japan, to assign geographical coordinates corresponding to the world geodetic system WGS 84 (World Geodetic System of 1984). The program identifies similar items in the addresses of the authors based on national format of addresses; it then isolates those that do not correspond to a national format of addresses which are then treated separately. Information resulting from this process is computed in the MapPoint geocoding engine to obtained standardized and localized information.

Table 1 shows how successful the process was since in the end we are in a position to geocode at least one address in 97% of articles and altogether 94% of all addresses identified.

Table 1 – Geocoding results of the dataset.

	Total	Successful geocoding	% of successful geocoding
Adresses	1 055 131	988 713	93,7%
Publications	538 074	520 914	96,8%

Localization of publications from 1998 to 2006 provides a first picture of science production which highlights its concentration. Three areas concentrate world production Europe (33%), North America (24%), Asia (33%), the rest of the world playing only a marginal role with 10% of the world knowledge production.

In these areas, five countries produce two thirds of all articles, giving testimony to the growing importance of China (but not India) and to the strong presence of France and Germany (but not UK) in Europe. Besides these main countries, 17 others are producing more than 1% of all publications i.e. 5500 publications or more (Table 2). As expected, the presence of countries from the triad is central since US, Japan and 10 European countries (including Switzerland and Israel) are in this list, as is also classically Australia. This table also demonstrates the presence of the three Asian tigers (Korea, Taiwan and Singapore), as well as the growing claim of BRIC countries on the world stage of scientific production (China is the third producer of articles; Russia is among the top 10 while Brazil and India are also there although to a lesser extent).

When considering the growth rates, this spatial distribution changes strongly and does not seem to limit itself to a “catching up” effect: we witness a strong growth is Asia with Japan as the exception. Thus, the map of world production capacities and potential is extending and evolving. Table 2 shows the changes in hierarchies that happen at the level of the top 22 countries producing at least 1% of world publications. Positions are relatively stable due to the advantage of initial critical mass existence[14] however China enters the top 5 from n.6 in 1998 to n. 2 in 2006. Taiwan progresses from n.16 to n.10. Despite the progresses in Asia, that is the most visible, we witness in less dynamic geographical areas countries taking ground such as Poland while other lose ground such as Sweden, country which has been

---

[14] Vincent Mangematin and Carole Rieu, nanotrendchart WP n° 2008-1, www.nanotrendchart.org

known for its agrofood and biotech businesses but that does not translate into a strong cluster when moving into the nano era.

Table 2 – Countries producing 1% or more of the knowledge production in nanoscience

Area	Country	Pub	Rank	TxVar	Pub_2006	Rank 2006	Pub_1998	Rank 1998
US & Canada	USA	134322	1	108,05	21718	1	10439	1
Asia	Japan	73136	2	71,72	10157	3	5915	2
Asia	China	70176	3	578,25	16339	2	2409	6
Europe	Germany	51409	4	69,49	7444	4	4392	3
Europe	France	35697	5	86,53	5400	5	2895	4
Europe	UK	31420	6	77,22	4620	7	2607	5
Asia	South Korea	27839	7	282,67	5254	6	1373	8
Other	Russia	19492	8	33,88	2351	13	1756	7
Europe	Italy	18799	9	123,58	2958	9	1323	9
Asia	India	16551	10	248,52	3412	8	979	12
Europe	Spain	15125	11	152,44	2643	11	1047	11
US & Canada	Canada	13368	12	134,06	2467	12	1054	10
Asia	Taiwan	13242	13	328,89	2865	10	668	16
Europe	Netherlands	9476	14	100,27	1462	16	730	13
Europe	Poland	8835	15	178,03	1468	15	528	18
Europe	Switzerland	8832	16	103,54	1439	17	707	14
Other	Australia	8754	17	156,18	1596	14	623	17
Europe	Sweden	8450	18	75,73	1202	20	684	15
Other	Brazil	8327	19	211,41	1392	18	447	20
Europe	Belgium	6747	20	119,96	1069	21	486	19
Asia	Singapore	6685	21	414,45	1317	19	256	22
Europe	Israel	5505	22	96,33	856	22	436	21

III. DATA GEOGRAPHICAL CLUSTERING

While looking at a world map through regional areas and national borders provides a first picture, a cluster analysis allows a better understanding of how scientific production spreads over countries.

Geographical clusters were constructed by selecting all cities that counted more than 1,000 addresses[15] i.e. each city being cited at least one thousand times in the database. These cities are called "core cities". There is a statistical tradition in the US to organize data into metropolitan areas but this does not apply in Europe and in Asia. To be consistent across continents, we thus developed a standardized routine in which all addresses that are located at a distance of 50 kilometers or less from core cities were brought together to constitute clusters[16]. Taiwan, South Korea and Japan are some exceptions to this: due to the human

---

[15] This threshold is relatively low considering that it represents an average of 111 addresses producing knowledge in the field per year over the period. This is especially low considering that we want to show general trends and detect places of knowledge production.

[16] We chose to build cluster of approximately 100 sq km considering that scientists can in this range benefit from cluster benefits described in the literature on districts (see for example David B. Audretsch et Maryann P. Feldman, Knowledge spillovers and the geography of innovation and production, American Economic Review, 1996, 86, 630-640.

We also used a radius of 30 kilometers to check the robustness of the 50 kilometer choice. This does not change the global trends described in this paper.

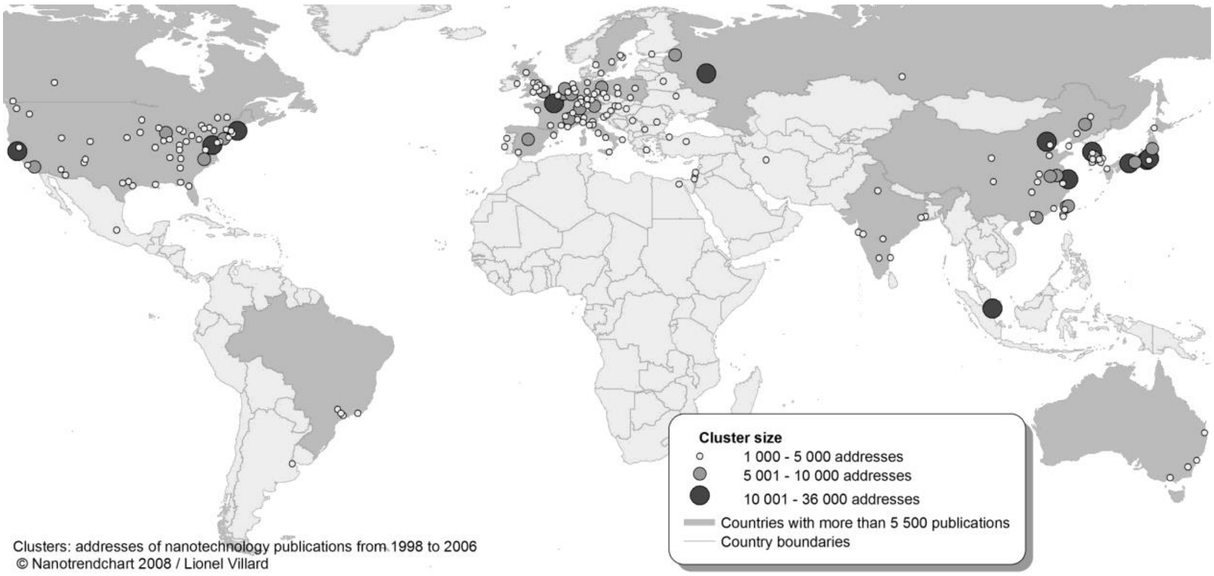
This method is automatic and due to the positioning of actors in near by cities, it is possible that smaller clusters are not integrated. This is, for instance, the case in France for the Bordeaux cluster: indeed, the cluster reaches concentration, we reduced the radius to 30 kilometers. This method brought us 293 initial clusters. However, some of them were overlapping: should they be regrouped into a larger cluster or not? Following statistical methods, we considered that if more than 20% of the addresses of the smaller of the two overlapping clusters were also present in the largest cluster, they should be brought together. At the end of the process, 200 clusters were generated (though round, the number is by chance!) regrouping 85% of all geocoded publications (444 558) and 73% of all geocoded addresses (722 066)[17].

Looking at cluster (Table 3) instead of nation level provides a different understanding of the world, especially taking into account their relative size (we have categorized them in 3 groups, small, medium and large), their dynamics (as translated by their rate of growth over the period) and their visibility (using the rate of citations of each publication per cluster).

Table 3 – localization of clusters by size and geographical area

Area	Nb of clusters	Small clusters (1000 to 5000 addresses	Medium size clusters (5000 to 10000 add)	Large clusters (more than 10000 add)
Asia	50	34	9	7
Europe	82	72	9	1
North America	50	43	4	3
Other	18	16	1	1

Figure 2 – Geographical distribution of clusters in the world



Clusters: addresses of nanotechnology publications from 1998 to 2006
© Nanotrendchart 2008 / Lionel Villard

---

more than 1000 addresses if we count Bordeaux, Pessac and Talence (the two latter being neighbouring cities in which some parts of Bordeaux university are settled in).

[17] A publication, with authors who have identified themselves at different addresses would be counted in several clusters. For discussions between fractioning versus full counting of publication, see for example A.J Nederhof and H.F. Moed. Modeling multinational publication: development of an on-line fractionation approach to measure national scientific output, Scientometrics, 1993, vol. 27, no1, pp. 39-52).

Shapira et al.[18], working only at US level have argued in their cluster analysis for a strong path dependency. Their central argument is that present clusters are strongly associated to previous high technology waves (ICT and biotechnology) or to the presence of a long-established powerful institution (university or national laboratory) that acts as an "anchor" to nanotechnology implication[19].

Taking a world view, we find quite similar results for the US but are driven to quite different conclusions for the rest of the world. Even in triadic countries we find "new" places, but the core of the differences lies not only in the clusters located in the new countries we mentioned, but also in places which have never been considered as important loci for previous technology waves. This is the case for European clusters located in areas previously under communist influence[20] Some other locations emerge in Latin America besides Brazil (Mexico and Buenos Aires) and in the Middle East besides Israel (Istanbul, Ankara, Cairo, Teheran). Thus, nanoscience production is indeed highly concentrated in a limited number of districts but it is also more diversified since areas not linked to previous technology waves are also present.

Indeed, during the period, twelve clusters have generated more than 10,000 addresses, and are present in 44% of publications (Table 4). This group has remained stable from 1998 to 2005, with only 2 new entries (Shanghai and Singapore), the 10 others already heading the list in 1998 (Tokyo, Kyoto, Tsukuba, Beijing and Seoul for Asia, San Francisco, Boston and Washington for the US, Paris and Moscow respectively for Europe and "Others").

Table 4 – rate of growth of the 12 large clusters

Area	Clust_Center_City	Nb_Add	Nb_Pub	Add_pub_98	Add_pub_05	Var	IGrowth
Asia	Tokyo	35363	25296	2928	4943	68,8	0,4
Asia	Beijing	26492	19692	990	5047	409,8	2,3
Asia	Kyoto	22285	16827	1809	3147	74	0,4
Asia	Seoul	20343	13529	964	3501	263,2	1,5
Europe	Paris	16385	11550	1416	2192	54,8	0,3
US & Canada	Berkeley	16176	11641	1294	2583	99,6	0,6
Asia	Tsukuba	14003	11159	647	2161	234	1,3
US & Canada	Washington	13292	9643	1003	2025	101,9	0,6
Asia	Shanghai	12347	9849	409	2533	519,3	2,9
US & Canada	Cambridge	11650	7887	815	1825	123,9	0,7
Other	Moscou	10368	7911	835	1339	60,4	0,3
Asia	Singapore	10256	6650	343	1795	423,3	2,4

Changes in hierarchy are even more visible when considering medium size clusters. Out of the 23 clusters that have between 1998 and 2006 gathered between 5,000 and 10,000 addresses, 2 belonged in 1998 to the top 12 (London and Berlin) and 4 were ranked beyond rank 40 : Hong Kong (40); Taipei (54); Heifei (67); Changchun (68). Moreover, out of the 23 clusters, the only ones that are better positioned in 2005 than in 1998 are Asian clusters with the most important progression being Hong Kong (from rank 40 to 16) and Tapei (from rank

---

[18] id. note 10

[19] Ajay Agrawal and Iain Cockburn. The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems," International Journal of Industrial Organization, 2003, vol. 21(9),1227-1253

[20] Krakow, Poznan, Warsaw and Wroclaw (Poland), Sofia (Bulgaria), Minsk (Belarus), Ljubljana (Slovenia), Bratislava (Slovakia), Kiev (Ukraine), Bucharest (Romania), Prague (Czech republic), Budapest (Hungary) and Belgrade (Serbia)

54 to 21). All 10 Asian clusters are progressing over the period in the hierarchy to the exception of Sendai and Nagoya. On the contrary, all 9 European clusters are losing ground most particularly London (from rank 11 to 27), Berlin (from rank 10 to 20), Zurich (from rank 13 to 22), Delft (from rank 18 to 29) and Munich (from rank 30 to 44). In North America, 2 clusters are stable, 1 is losing ground and 1 is gaining ground. Last, St Petersburg is losing the most ground out of all medium size clusters going over the period from rank 20 to rank 41. This tendency is also visible for smaller size clusters (Table 5).

Table 5 – Progression, stability and loss of ground in hierarchy by size of cluster and geographical area

Area	Cluster size								
	Small			medium			Large		
	progress	stable	lose ground	progress	stable	lose ground	progress	stable	lose ground
Asia	26	1	6	7		2	5	2	
Europe	8	15	48			9			1
North America	17	10	21	1	2	1		2	1
Other	6	2	8			1			1

In the econometric model of Mangematin and Rieu[21], these two aspects, the continental localization and the initial size of clusters explain more than 60% of cluster growth. Between 1998 and 2005[22], the average growth of clusters is 11% per year and is borne by one fourth of clusters. Out of these 50 clusters[23], 33 are in Asia (6 in Europe, 5 in the US and 6 in Other). One major result is however, that initial positions matter

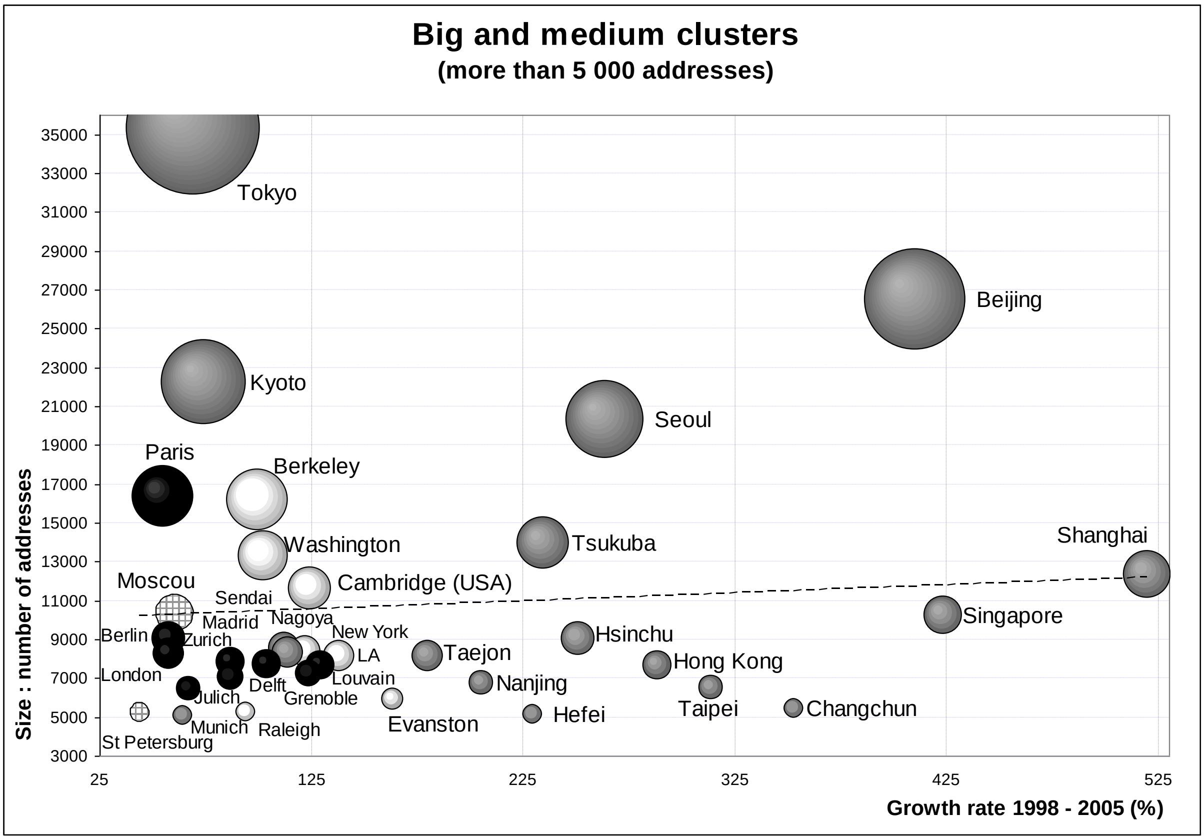
---

[21] Reference cited note 14

[22] 2006 is not included since 2006 extraction was not performed on a full year basis.

[23] Out of the 50 clusters, 5 are large ones, 6 are medium size ones and 39 are small clusters.

Figure 3 – Rate of growth of large and medium size clusters depending on their size and geographical area



Note: the size of the sphere is proportional to the size of the cluster. Geographical areas are represented as follows: black for Europe; gray for Asia; white for North America; white with gray stripes: other.

IV. CHARACTERIZING CLUSTERS : VISIBILITY AND THEMATIC SPECIALIZATION

Rate of growth does not allow encompassing all the diversity of clusters. This is just one element to consider. Considering the role of size that others[24] have pointed on clusters, we thus develop two hypotheses in relation to it.

HYPOTHESIS 1 : THE VISIBILITY OF A CLUSTER IS LINKED TO ITS SIZE

Clusters can grow more or less fast in publications but it does not tell anything about the quality of the production. For each cluster, we measure it by adding all citations received by publications counted in a cluster. We then establish a ranking on clusters having the larger proportion of citations in the top 1% of publications the most cited and in the top 0.1% of the publications the most cited. This measures the visibility and market share of clusters in the new knowledge production industry.

31 clusters have publications included in the top 0.1% of cited publications, which represent 98 publications (Table 6). More than half of them (17 out of 31) count only one publication on the period, while 6 clusters (all except one are in the USA) account for more than 50% of the most cited publications. 14 of the clusters are located in the USA accounting for 62% of

---

[24] Mangematin and Rieu cited note 14; Porter et al, cited note 10

the most cited publications. Asia counts only 6 clusters (9 highly cited publications) and Europe, 13 clusters and 27 highly cited publications. The size of the cluster is not the prime determinant of the quality of publications as Houston (5), Santa Barbara (8) or Atlanta (5) that are small clusters, count more highly cited publications than large clusters such as Tokyo, Kyoto, Paris or Baltimore (1 publication each).

The picture is the same when considering clusters contributing to the top 1% cited publications (9381 publications). 174 of the 200 clusters contribute to the top 1% by less than 1%. Again we see that it is not only a question of size, as 3 are large size clusters (each contributing to 0.3% of the top 1% cited publications) and 14 are medium size clusters. Not surprisingly, the top contributors are those already contributing to the top 0.1% with a few notable exceptions.

Table 6 – most visible clusters (top 0, 1% and top 1%)

Clust_Country	Clust_City	Nb_add	Nb_pub	Pub_In_Top 0.1%	Contribution to the top 0.1%	Pub_In_Top 1%	Contribution to the top 1%
USA	San Francisco	16176	11641	14	14%	654	7%
USA	Boston	11650	7887	14	14%	609	6%
USA	Baltimore	13292	9643	1	1%	343	4%
Japan	Tokyo	35363	25296	1	1%	286	3%
USA	New York	8418	6177	2	2%	262	3%
USA	Los Angeles	8142	5973			261	3%
USA	Houston	4090	2619	5	5%	231	2%
USA	Chicago	5963	4186	1	1%	212	2%
France	Paris	16385	11550	1	1%	193	2%
Switzerland	Zurich	7847	6284	1	1%	190	2%
USA	Santa Barbara	3536	2453	8	8%	188	2%
Japan	Kyoto	22285	16827	1	1%	177	2%
USA	Chapel Hill	5288	3847			163	2%
Netherlands	Amsterdam	7059	5264	5	5%	147	2%
USA	Philadelphia	3885	2757			145	2%
Germany	Berlin	9065	7662	2	2%	140	1%
England	London	8270	6720	3	3%	120	1%
England	Cambridge	4973	4259	3	3%	113	1%
Germany	Munich	5106	4057			107	1%
South Korea	Seoul	20343	13529	3	3%	100	1%
Israel	Jerusalem	4841	3532			99	1%
China	Beijing	26492	19692			99	1%
Japan	Tsukuba	14003	11159			96	1%
USA	Pittsburgh	3518	2625			92	1%
USA	Ann Arbor	4011	2830			92	1%
USA	Atlanta	3976	3059	5	5%	91	1%

The landscape is quite different when considering publications in the top 10% of the most cited publications. Clusters that have the largest share of these are all American clusters, Philadelphia, Santa Barbara and Boston having more than 30% of their publications in the top 10%. Out of the first 20, we find only Jerusalem and Groningen.
Compared to Table 6, Paris falls at the 62nd position, Tokyo at the 109th place and Zurich at the 35th position.

When dealing with the position of clusters in the top 10%, size is definitively not a good indicator: out of the 35 clusters having more than 20% of their publications in the top 10%, 25 are small ones, 7 medium size clusters and only 3 are large ones

So, in relation to hypothesis 1, we can say the visibility of a cluster here indicated by the quality of its publications in the top 0.1% and 1% of the most cited publications, is not related to its size

HYPOTHESIS 2A : THE THEMATIC SPECIALIZATION OF A CLUSTER IS LINKED TO ITS SIZE

HYPOTHESIS 2B : THE SIZE OF A CLUSTER IS LINKED TO ITS MARKET SHARE IN THE FIELD

A further assumption about concentration in a limited number of clusters and about their very different growth lies in their thematic specialisation. Our assumption is that the strength (real or potential) of a cluster relies on the 'convergence'[25] of previously separate disciplines associated respectively with electronics, biotechnology and materials. In order to make a first test about this assumption, we have organised data around three main aggregates[26] linked to (1) electronics and physics, (2) materials and chemistry, and (3) nanobiology and life sciences.

A first simple count can be made in relation to clusters' respective global shares and share evolution over the period (1998-2005). It shows that nanobiology remains quasi stable at a low level (11% in 1998 vs 12% in 2005). But there is a significant shift between the two other aggregates: the growth of nanomaterials is extremely fast, driving it share up from 47 in 1998 to 53% in 2005 (an increase in nearly 1% every year) while this is nearly the opposite for nanoelectronics.

However, a closer attention is required. As shown in Table 7 and 8, there are wide variations around the average with very strong specialisations both in relative shares (Table 5) and in publication production (Table 8) : some clusters having, for instance, more than 80% of their publications in materials and chemistry (e.g Xiamen and Harbin, China) , and other clusters having less than 35% of their production in biology (e.g. Philadelphia, PA; St Louis; MO, USA).

Table 7 - Cluster specialisation: respective shares of clusters

%	Average	minimum	maximum
Nanomaterials	50,1	29,2	83,4
Nanoelectronics	35,2	11,9	66,8
Nanobiology	11,6	1,1	40,7

Table 8 - Cluster specialisation: respective sizes of production*

publications	average	median	minimum	maximum

---

[25] see for example Rafols, 2007

[26] In a recent 10-year period Thomson Scientific recorded about 9 million articles, notes, and reviews, published in roughly 9,000 indexed journals. These are categorized in 9 or 123 categories by JRC ISI. Journals can be assigned to one or more categories, which makes the classification more blur. We preferred the *Essential Science Indicators* from In-Cites that categorizes these journals into 22 broad disciplines. "Each journal is assigned to one and one only of the 22 disciplines. Similarly, *Essential Science Indicators* then assigns each paper to a discipline—and only one discipline—based on the journal in which it appears. In the case of multidisciplinary journals, special processing is carried out to assign individual papers to fields based on the predominate field of the papers' citations and references" (www.In-cites.com). We then re-assigned these 22 categories into 3 large fields that are more easily understood by experts in nanosciences. See appendix A for more details.

Nanomaterials	1905	1298	381	15922
Nanoelectronics	1440	809	130	15719
Nanobiology	441	261	15	3347

* production is measured by the number of addresses in publications

This data must however be read with caution, when account is taken of the very skewed distribution of clusters by size. As mentioned above, the 35 largest clusters together account for nearly 50% of addresses that are organised at 25% in nanobio, and 53-54% for the two largest fields, nanoelectronics and nanomaterials. (Table 9).

Table 9 - Medium and large size clusters specialisation (%)

clusters	Share global	nanomaterials	nanoelectronics	nanobiology
12 largest	27%	29,6	29,3	23,6
23 medium	21.5%	23,3	22,5	21,6

In order to control for this effect, we have complemented internal specialisation ratios with roles played by clusters in each domain. We thus build a measure of cluster involvement considering that it could offer a hierarchy within specialised areas that relative shares of size individually cannot grasp. We calculate it in the following way: when its effective size (in term of addresses) is over 25% of the average one, we attribute it a higher than average role (noted H); when it is twice as large as the average one, we attribute it a very high role (noted VH). On the contrary, when size of a cluster is 25% lower than average, we consider its role to be below average (noted B); when it is twice as small as the average cluster size, we consider its role to be very small compared to the average (noted VS) (Table 10).

Table 10 – characterisation of the role clusters in their domain (thresholds)

	Degree of specialisation (number of addresses)				
	VH	H	A	B	VS
Nanomaterials	> 3810	> 2381 and < 3810	1905	< 1905 and >1428	< 953
Nanoelectronics	> 2840	> 1800 and < 2840	1440	< 1440 and > 1080	< 720
Nanobiology	> 882	> 551 and < 882	441	< 441 and > 331	< 220

What we see first is that the relationships between the internal specialisation of a cluster and its role in the domain is first an issue for smaller clusters, which even if very highly specialised have difficulty to play even an average role in their domain of specialisation. On the contrary, whatever their degree of specialisation, the 12 largest clusters always play a very large role in both electronics and materials, while the situation is more contrasted in biology where a ‘very high role’ (meaning some 880 addresses) can be attained with ‘relative’ specialisation by ‘big’ small clusters (between 2500 and 5000 addresses). Similarly, except in biology, where only 13 out the 23 mid-size clusters play at least a high role, we find 18 in materials and 21 in electronics playing a high or very high role.

A striking element is that specialisation is largely linked to history and materials and chemistry is highly important in the development of nanoscience. Asia is very strong in materials and electronics: we see within this large area differences between the countries known as the dragons that have been focusing on large markets for electronics and China which has the largest specialisation in materials. Table 11 shows this for mid size clusters but this is applicable to other ones. On the opposite, the only places focused on biology are American ones (even Oxford is invisible in terms of role in the field), where biotech emerged. In Europe, path dependency is strong

Table 11 – Role and specialisations of medium size clusters

	Role in the field			Publication share		
	Nanobio role	Electronics role	Materials role	Nanobio special	Electronics special	Materials special
Berlin	VH	VH	VH	A	**S**	A
London	VH	H	H	**VS**	A	**B**
Zurich	VH	H	H	**S**	**B**	A
Madrid	H	VH	VH	B	A	A
Louvain	VH	H	H	**S**	A	A
Grenoble	A	VH	H	**FB**	**VS**	A
Delft	VH	H	H	**VS**	A	A
Julich	H	H	H	A	**S**	A
Munich	VH	H	A	S	**S**	FB
Hsinchu	L	VH	VH	FB	VS	A
Nagoya	A	VH	VH	B	A	S
Sendai	L	VH	VH	FB	VS	A
Taejon	L	VH	VH	FB	S	S
HK	A	VH	VH	FB	S	A
Nanjing	L	H	H	FB	S	A
Taipei	H	H	H	A	S	A
Changchun	VL	A	VH	FB	FB	VS
Hefei	VL	A	H	FB	B	VS
NY	VH	H	H	S	L	B
LA	VH	H	H	S	L	B
Chicago	VH	A	H	S	FB	A
Raleigh	VH	A	A	VS	L	B
St Petersburg	VL	VH	A	FB	VS	FB

In relation to hypothesis 2a and 2b, we show that small clusters tend to be more specialised than larger clusters. However, we also demonstrate that a high degree of specialisation is not an indicator of a dominant position in the field: large clusters do not have to be much specialised in a field to play an important role in the production of knowledge in nanoscience. On the contrary, small clusters that are more often highly specialised have difficulty playing even an average role in their field. Note that in biology, as the threshold is lower, smaller clusters can pretend playing a role.

V. CONCLUSIONS

What does data tell us?

The very first element that this paper documents is the high concentration of forces in the production of knowledge in nanosciences: the world is not flat but many hills punctuate the landscape. However, we also demonstrate that not all hills are alike: some are rather mountains, while many smaller hills exist. Looking at the rate of growth of clusters: interestingly, the most dynamic clusters are all but 2 located in Asia; Europe and Americas being far being Asia. However, the most visible clusters are all but 2 located in Europe and in the US. Interestingly the height of mountains is not linked to their visibility as we have shown that small hills are very visible.

We show as we detail the role of clusters and their rate of growth, the necessity to create critical mass in order to produce new knowledge. This statement relies on theories on geographical agglomeration advantages such as the access to a qualified labour force, to suppliers and to the presence of "something in the air"[27] etc. It also related to the characteristics of knowledge in emerging fields: knowledge is difficult to grasp and geographical proximity makes it more easily transferable[28]. Physical proximity also allows spillovers due to more easily day-to-day interactions between actors[29]. Out of the 200 geographical areas, the 12 large size ones and the 23 medium size ones account for 47% of knowledge production. The critical mass is more easily achieved in areas that already had competences in biology, physics, electronics or informatics from past technological wave. But we also show that new places, in or outside of the triadic countries emerge. Moreover, the quality of the production is not only dependent on this sole criterion of mass effect as very small clusters manage to have a relatively large market share of very high quality publications.

The extent to which a cluster can benefit from critical mass effects is also strongly history dependent: biology is only important in the US, where biotech started; electronics is strong in the four dragons as well as in Europe, while materials are dominated by China. As we link this with the growth of rate of clusters, we show more general trend; materials and chemistry are growing and are more and more important to the expense of electronics essentially, while biology plays a role in only a few places in the world. Thus the landscape punctuated by hills becomes more and more coloured by materials and by a few mountains and many small hills. However, it is constantly evolving as places appear (like in Eastern Europe or in India) while some become less important despite the bygone splendor.

However, the extent to which the belonging to a specific country would an indicator of rate of growth, size or visibility, is very limited. The diversity of clusters illustrates this. This has strong consequences for public policy design which thus might not be relevant at the country

---

[27] Marshall, Alfred (1920). Principles of Economics, Revised Edition, London: Macmillan; reprinted by Prometheus Books

[28] Nonaka, Ikujiro & Takeuchi, Hirotaka (1995). The Knowledge Creating Company. New York: Oxford University Press

Polanyi, Michael. The Tacit Dimension. First published Doubleday & Co, 1966. Reprinted Peter Smith, Gloucester, Mass, 1983. Chapter 1: "Tacit Knowing"

[29] Jaffe, Adam B & Trajtenberg, Manuel & Henderson, Rebecca, 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," The Quarterly Journal of Economics, MIT Press, vol. 108(3), pages 577-98, August

Anselin, Luc & Varga, Attila & Acs, Zoltan, 1997. "Local Geographic Spillovers between University Research and High Technology Innovations," Journal of Urban Economics, Elsevier, vol. 42(3), pages 422-448, November.

Audretsch, David B & Feldman, Maryann P, 1996. "R&D Spillovers and the Geography of Innovation and Production," American Economic Review, American Economic Association, vol. 86(3), pages 630-40, June.

Feldman, Maryann P. & Audretsch, David B., 1999. "Innovation in cities:: Science-based diversity, specialization and localized competition," European Economic Review, Elsevier, vol. 43(2), pages 409-429, February.

level: a single national policy is not coherent to encompass the diversity of clusters: the most valuable example of this limit is the existence of cross-border clusters.

However, to consider policies based on the size and the degree of specialisation of clusters may be more relevant: we have indeed shown that both these elements need to be consider to understand the importance of a cluster in a field compared to the other ones. Large clusters are always central players in almost the 3 fields while medium sized ones are important to very important in 2 fields. On the contrary small clusters have difficulties, even if specialised to reach an average position in the field.

Appendix

Incites 22 categories and ventilation of publications from our database

InCites 22 categories	Nb_Nano_publications
Chemistry	146260
Materials Science	140748
Physics	138418
Engineering	41106
Biology & Biochemistry	17037
Clinical Medicine	15080
Pharmacology & Toxicology	10673
Geosciences	5373
Molecular Biology & Genetics	4444
Plant & Animal Science	4071
Multidisciplinary	3829
Environment/Ecology	2936
Agricultural Sciences	2310
Microbiology	2222
Neuroscience & Behavior	2140
Computer Science	1749
Space Science	1276
Immunology	1194
Mathematics	896
Social Sciences, general	669
Economics & Business	419
Psychiatry/Psychology	181

From In-cites 22 categorie classification to our 3 category classification

Field	**Field_Nano**
Physics	NanoElect/Phys
Chemistry	NanoMat/Chem
Materials Science	NanoMat/Chem
Engineering	NanoElect/Phys
Biology & Biochemistry	NanoBio
Clinical Medicine	NanoBio
Pharmacology & Toxicology	NanoBio
Molecular Biology & Genetics	NanoBio
Multidisciplinary	Other
Geosciences	Other

From In-cites 22 categorie classification to our 3 category classification

Field	Field_Nano
Neuroscience & Behavior	NanoBio
Plant & Animal Science	NanoBio
Microbiology	NanoBio
Environment/Ecology	Other
Space Science	Other
Immunology	NanoBio
Computer Science	NanoElect/Phys
Agricultural Sciences	NanoBio
Mathematics	Other
Social Sciences, general	Other
Economics & Business	Other
Psychiatry/Psychology	Other