December 2, 2003

VOLUME 5 (2001): ISSUE 1. PAPER 2

Indicators of Innovation in a Knowledge-based Economy

Loet Leydesdorff

Science & Technology Dynamics, University of Amsterdam
Department of Communication Studies, Oude Hoogstraat 24
1012 CE Amsterdam, The Netherlands
loet@leydesdorff.net
http://www.leydesdorff.net

Abstract

The concept of 'modes of knowledge production' was used by Gibbons et al. (1994) [1] to distinguish between transdisciplinary ('Mode 2') R&D and more traditional ('Mode 1') research. This paper explores whether the Internet provides a means to operationalize 'Mode 2' knowledge production as containing a differently codified communication pattern which can be compared to co-word and citation patterns in scientometric databases ('Mode 1'). Innovations on the drugs market, for example, can be indicated at the commercial end by using the trade names of the drugs (e.g., *Evista*), while the very same innovation can be retrieved in the patent and science citation databases using the generic names of the active substances involved (in this case, *raloxifene*). By using the generic names the new drugs can be traced back into their respective knowledge bases.

Keywords

innovation, pharma, scientometrics, citation, search, knowledge

1. Introduction

In a knowledge-based economy, information exchange processes provide a mechanism of social coordination in addition to economic exchange relations and political and/or managerial control. The mechanisms of information transfer can be operationalized in terms of the measurement of communications at relevant interfaces. In general, communication systems develop interactively in terms of networks, but also recursively in relation to their previous stage. The recursively embedded codifications can be changed by the interactions. Knowledge-based innovation systems thus provide us with a system of social coordination in terms of communication that is potentially coded differently in scientific and market domains.[2],[3]

In this study, I focus on knowledge-based innovation in pharmaceuticals. On the one side, applicational contexts (e.g., public health issues, user demand, and patient organizations) can be expected to structure the communication at the commercial end. This horizontal codification may enable us to

operationalize the dimension of applicational contexts, or the 'Mode 2' of scientific knowledge production,[1] in terms of an empirically retrievable, yet independently measurable interfaces, for example, at the Internet. The products of science and technology are interfaced with the public both on the market and with reference to laws and regulations. On the other side, the knowledge base of these innovations ('Mode 1') can be measured using scientometric methods. The Science Citation Index is highly codified, privately owned, and its accessibility is controlled. Scientific citation patterns primarily reflect the vertical organization of knowledge codification within the sciences. [4]

Can patents and patent citation databases be considered as a linking pin between these two modes of knowledge production? Patent databases and indicators are codified to a variable extent, that is, in relation to their envisaged usages. For example, the databases of the European Patent Office and the U.S. Patent and Trademark Office are publicly available at the Internet, but on the side of the knowledge-base, the systematic organization of patent citations into a data base can be considered as added value.[5],[6], [7],[8] The Derwent Patent Citation Index is, therefore, commercial. [i]

In general, data is available at the Internet for the period since 1993.[ii] This seems sufficient data for the analysis of recent technological innovations. Evista, for example, was introduced in 1997 by Eli Lilly as a new drug against osteoporosis. However, it is based on patents from the late 1970s, owned by this same corporation.[9] I will compare Evista with the drug Fosamax brought on the market in 1995 by Eli Lilly's competitor MSD, on the one hand, and with Prozac, an older (1988) drug of Eli Lilly, on the other. My research question is mainly methodological: is it possible to map an innovation using the various databases as indicators of codification along assumingly different axes? Which are the relevant dimensions and what may one expect if one wishes to upscale this type of analysis, that is, with a focus on innovation as the unit of analysis? [10],[11],[12]

In addition to patent data available on the Internet, I will use the (codified) Derwent Patent Citation Index for the systematic retrieval of their knowledge base. Medline is used for the delineation of the medical domains of application. The Science Citation Index remains a crucial instrument because of the systematic efforts to standardize corporate addresses in addition to providing intellectual lineages in terms of references. The Derwent patent citation database can be considered as its systematic pendant in the patent literature. By limiting searches at the Internet to national language domains one can additionally find windows on national agencies like patient organization[13]. This is not the case for Fosamax or Evista .

2. Theoretical Context

Previous studies have taught us that the patterns in the relations between patent literature and scientific literature can be counterintuitive.[14],[15] The Internet additionally allows for measuring informal communication[16]. In general, indicators provide us with a methodology for the measurement within their respective domains of application, since indicators are usually developed with reference to a specific domain.

Does the neo-evolutionary model enable us to specify a framework for the appreciation of the interaction? University-industry-government relations reshape the institutional carriers and their interactions in innovation systems. [17] Each interacting sphere develops both recursively and by its network of interactions. The systems of knowledge production, diffusion, and control feed back on each other as subdynamics of communication, while developing continuously along their respective trajectories. The various interacting subdynamics recursively shape each other and the interactions can be expected to result in a complex dynamics[18],[19]. Each axis is guided by its own code of communication, but during the operation the codes may also

change due to events taking place at the interfaces.[20]

One can expect science-based innovations when knowledge is **(i)** recursively generated, **(ii)** sufficiently codified (before utilization) so that it can be formally transmitted, and **(iii)** the subject of market forces. The interfaces can be expected to 'lock-in'[21] over time as in processes of 'mutual shaping.'[22] To the extent that the complexity of the system under study evolves and functions can further be differentiated, the various interfaces can also be expected to develop over time into different functions [23],[24]. The databases tend to serve distinguishable audiences and routines, Some databases are dedicated (e.g., Medline), while others (e.g., the Science Citation Index) have remained (more) general purpose databases.

3. A stylized history of Evista

The starting point of our research was provided by Mogee & Kolar's (1999) [10] study of the patent co-citation relations among Eli Lilly patents. Using the 2808 U.S. patents of Eli Lilly issued from 1975 through to 1998, these authors concluded that the patents leading to the development of Evista compose a key cluster (G11C5) in the co-citation patterns of this firm (Figure 1).
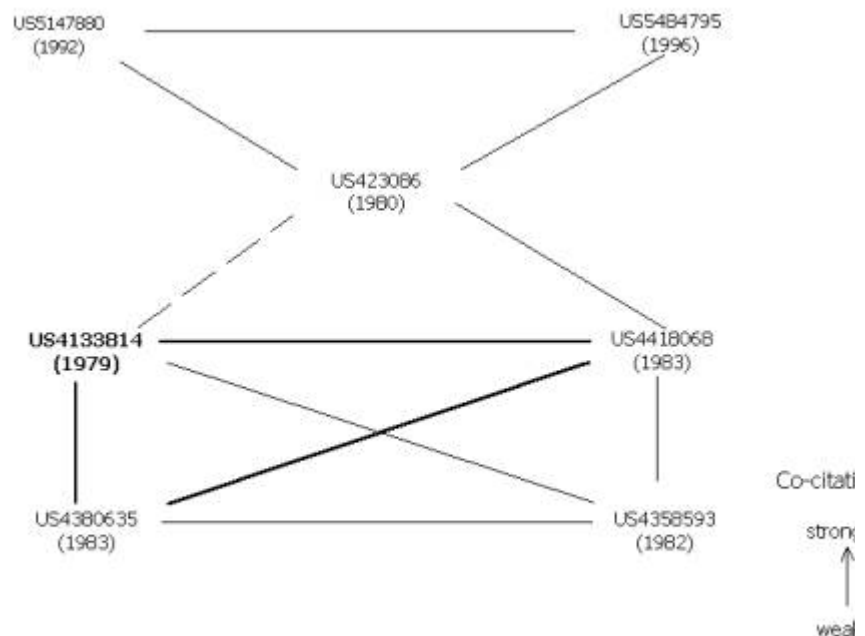
Figure 1: Diagram of Cluster G11C5, from: Mogee & Kolar (1999,[10] at p. 302) publication dates added.

These patents cover the preparation of the compounds and methods of using them to treat a range of conditions normally treated by oestrogen. The advantage of these compounds is that they lack the negative side-effects of oestrogen (Mogee,[10] at p. 301). The seven patents were heavily co-cited by other patents of this corporation in the period around the introduction of Evista in 1997: 173 patents cite into this cluster of which 139 citations come in the years 1996-1998.[25] Note that all citations are 'in-house,' that is, within the corporation.

The original patent in this series is labeled as US4133814 from 1979 (boldface added, L.). The new compounds are described in this patent as antifertility agents that might also be useful in suppressing the growth of mammary tumors. Patent US4418068 includes, among other compounds, raloxifene, the

active ingredient of the 1997 drug Evista™. Note that this patent dates from 1983. These compounds were at that time mainly considered as useful for anti-oestrogen therapy in the case of mammary tumours.

The series of five patents between 1979 and 1983 was continued with US5147880 in 1992 and US5484795 in 1996. The 1992 patent describes the oestrogenic and anti-oestrogenic activity of these same compounds. Among the therapeutic applications is now mentioned the prevention of bone loss. The 1996-patent relates to a different group of non-steroidal compounds that are useful for the treatment of medical indications associated with the post-menopausal syndrome.

In 1997, Lilly filed for approval of Evista by the Food and Drug Administration (FDA) of the U.S.A. Evista can be used for the prevention of osteoporosis in post-menopausal women. The drug itself had been in clinical development for more than five years at this point in time. Sales were at US$ 130 million in the year of launching the drug (1998) and they are expected to go steeply up to over one billion dollars in the near future. In summary, a group of patents from the early 1980s provided Eli Lilly with the knowledge base to develop Evista in the 1990s.

The seven patents were cited in the Science Citation Index 14 times since 1988, [iii] including eight times in 1997 and twice in 1993. Citations, however, were exclusive to the earlier group of patents; both the 1992- and the 1996-patent had zero citations (in November 1999). Nine of the 14 citations were of the original patent US 4133814, including four in 1997. Half of these citations are in journals in the area of 'bioorganic and medical chemistry,' although there is also citation activity (4 times) in a clinical journal entitled Expert Opinion on Therapeutic Patents. The remainder appeared in pharmaceutical journals.

It seems from this data that there is relevant research activity in following the applications of these compounds in therapies. For example, S. Murthy and A. Flannigan cited U.S.-patent 4230862 of 1980 in their article entitled 'Recent developments in inflammatory bowel disease therapy,' (Epert Opin. Ther. Pat. 7 (1997), 695-715). In summary, the research front makes references to the groups of patents including raloxifene—the generic name of Evista—but one cites predominantly the original patent of the group.

4. Methods and Materials

The searches reported here below were performed during the period of November/December 1999 (unless otherwise indicated). Table 1 provides the keywords used as search terms and summarizes some relevant information concerning the drugs under study.

Drug (tradename)	Generic name	Applicant	Publication date of patent
Evista	raloxifene	Eli Lilly	1997
Fosamax	alendronate	MSD	1995
Prozac	fluoxetine	Eli Lilly	1988

Table 1. The three drugs compared

Fosamax™ is the brand name for alendronate which was patented by Merck & Co (MSD), one of Lilly's competitors. The Food and Drug Administration (FDA) approved Fosamax as another new drug to treat osteoporosis on October 3, 1995. From a user perspective, the two products can be considered as alternatives[iv]. The other comparison will be with Prozac™, a well-known drug brought on the market by Eli Lilly since 1988. Prozac was the world's

best-selling anti-depressing during the 1990s. It is the brand name for fluoxetine which was patented by Eli Lilly in 1988, that is, before the introduction of Internet browsers around 1992. Prozac provides us with a baseline because it can no longer be considered as an innovation in the second half of the 1990s.

The keywords indicating the trade names and the generic names for Evista, Fosamax, and Prozac were compared in terms of the retrieval using the Science Citation Index, the Derwent Patent Citation Index, Medline, and the AltaVista Advanced Search engine at the Internet. The AltaVista Advanced Search Engine was used because it provides superior capacities combining date delimiters with a full set of Boolean operators on various possible tags (e.g., domain, link, and language). **[26]**

There are some severe limitations with using web data and search engines. First, dates stamped to webpages mean something different from publication dates.**[27]** Webpages begin their lifecycle at the date of their publication, while a scientific publication is final at the date of its appearance. Webpages can be updated and then changed in terms of their publication dates and the webcrawler can also add pages with hindsight and at potentially earlier dates. Thus, the Internet system is evolving not only along the time axis, but also in the present (i.e., with hindsight).

Second, the Internet can be considered as an emerging phenomenon that is a (virtual) result of interactions among more specific representations. From this perspective, the Internet itself remains a hypothetical domain to be mapped by using one or more search engines. Reflexively, one can, for example, study the quality of the search engine, for example, by making comparisons among them**[28]**. The various search engines can be expected to provide us with very different results because they use different angles (and algorithms) for the representation.

Given these limitations in using web data, the quality of the organization of the data within each representation provides an important criterion for using one search engine or another. (The combination of search engines tends to confuse further the methodological control.) Most search engines are user-oriented, but some of them provide analytical tools in so-called "advanced" versions. Among the major search engines, the AltaVista Advanced Search Engine provides hitherto superior capacities for combining both date delimiters and a full set of Boolean operators. **[v]**

5. Mapping the competing drugs at the Internet

Using Altavista's advanced search options,**[26]**,**[27]** we found the following time series for the four (two commercial and two generic) names for the two competing drugs (Fosamax and Evista):
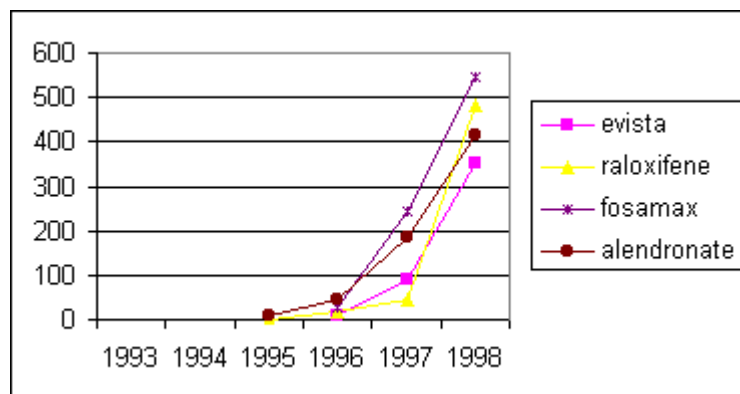

Figure 2. Number of hits using AltaVista's Advanced Search Engine for both

generic and trade name of two competing drugs

The graphs show the somewhat lagged competition of Evista with Fosamax. Although Evista was introduced two years later than Fosamax, 'raloxifene' had grown slightly more visible in 1998 than 'alendronate.'

In Figure 3, these time series are extended with the analogon for the longer existing drug Prozac™. The curve for 'Prozac' versus its generic name 'fluoxetine' informs us that the commercial brandname, indeed, is more important on the Internet than the generic name when a drug is widely known and increasingly used **[vi]**. This is not yet the case for Fosamax or Evista.
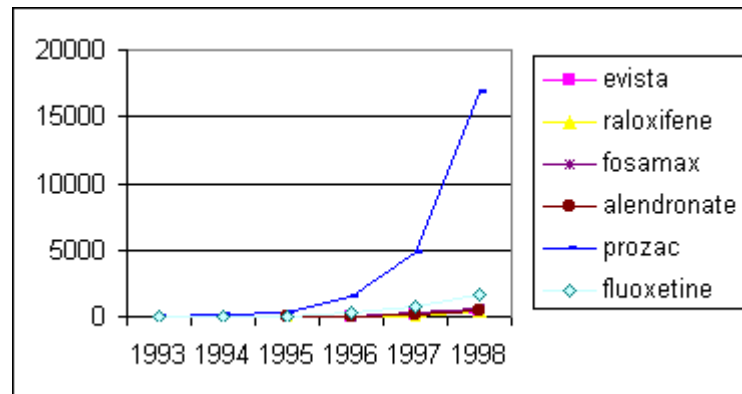
Figure 3. Size effect of number of hits in the case of a widely marketed product

6. Mapping the knowledge infrastructure at the Internet

The search engine enables us also to constrain the searches to specific domains. In addition to national domains (e.g., '.fr' for France), the so-called 'generic Top Level Domains' (gTLDs) can be decomposed in terms of the functions of these domains like '.com' for commercial, '.edu' for educational, etc. One expects that the '.edu' domain indicates mainly the university-side of the knowledge interface, while the '.com' domain is dominated by the commercial side. (The '.gov'-domain refers exclusively to agencies of the U.S. government.)

This indication of institutional domains can then be cross-tabled with the generic versus the trade names. In both cases, the cross-tabulated distributions for 1998 are significantly different (p < 0.01): the generic name 'raloxifene' is more often retrieved than 'Evista,' even on the '.com' side, while the commercial name 'Fosamax' is more common than 'alendronate' even within the '.edu' domain. How may this indicate a different relation to the knowledge base in either case?

a. trade names

Prozac and Evista as trade names are predominantly present within the '.com'-domain, while Fosamax seems to a large extent also still a subject of academic discussion in the '.edu'-domain. The difference can be shown by the comparison of figures 4a and 4b:
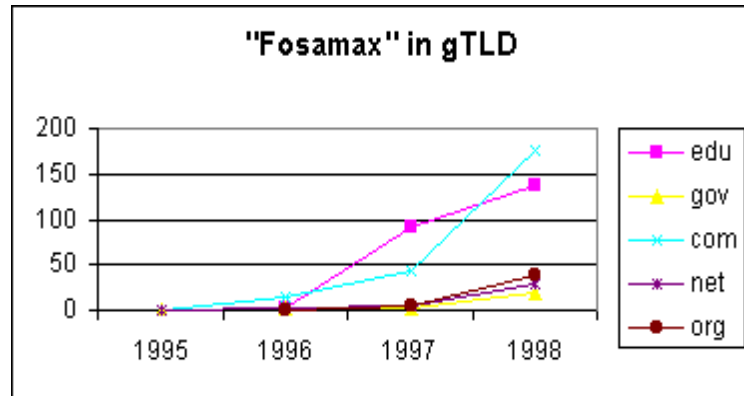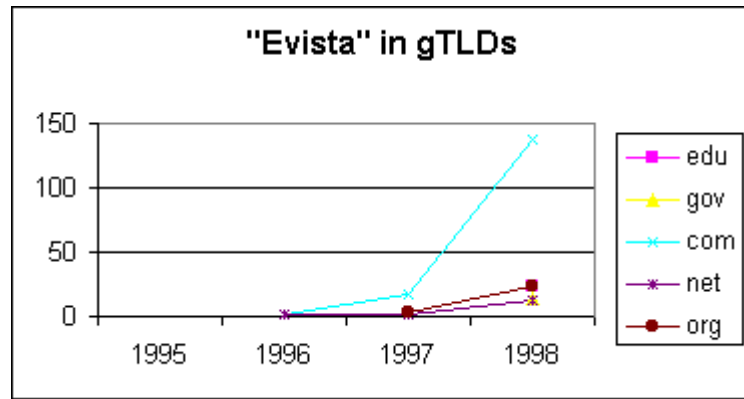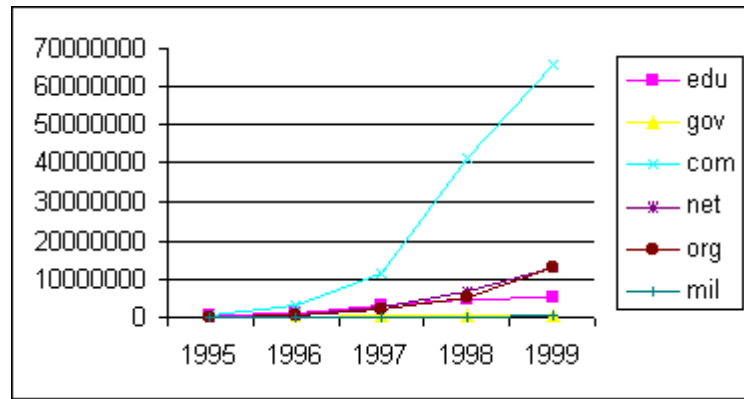
Figures 4a & 4b. Trade names differentiated among different domains

However, the relative contributions of domains to the Internet changed significantly from the first to the second half of the 1990s, as shown in Figures 5a and 5b.
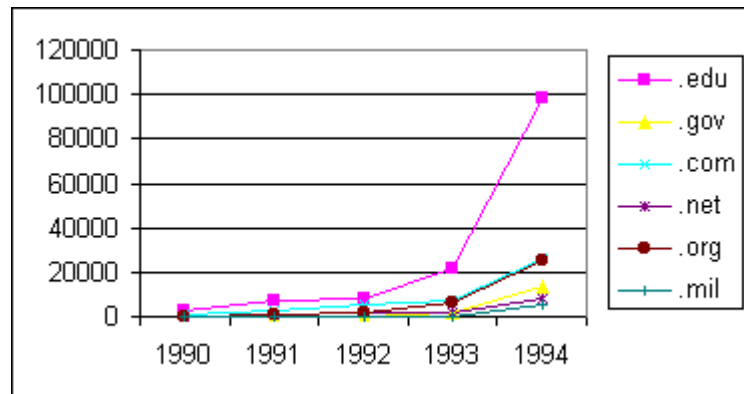
Figure 5a. Differentiation among generic Top Level Domains 1990-1994 (from: Leydesdorff, 2000).[29]

Figure 5b. Differentiation among generic Top Level Domains 1995-1999
(from: Leydesdorff, 2000).**[29]**

Note that the difference in the scales of the y-axis between figures 5a and 5b
is in terms of orders of magnitude. In the first half of the decade, the
educational sector was booming at the Internet, while the commercial sector
('.com') took the lead during the second half. Commercial domination like in
the case of Evista and Prozac (Figure 3) can, therefore, be expected in this
period. From this perspective, the visibility of Fosamax in the '.edu' domain
indicates an exception which requires further explanation.

b. generic names

Figures 6a, 6b, and 6c provide the patterns for the generic names of these
three drugs in terms of the same decomposition. In this case, the prevailing
predominance of the commercial sector is relatively suppressed in favour of
mainly the domains '.edu' and '.org,' providing a clear indication of the
respective relations to the knowledge infrastructure. The generic names have
remained relevant in domains other than the commercial ones.
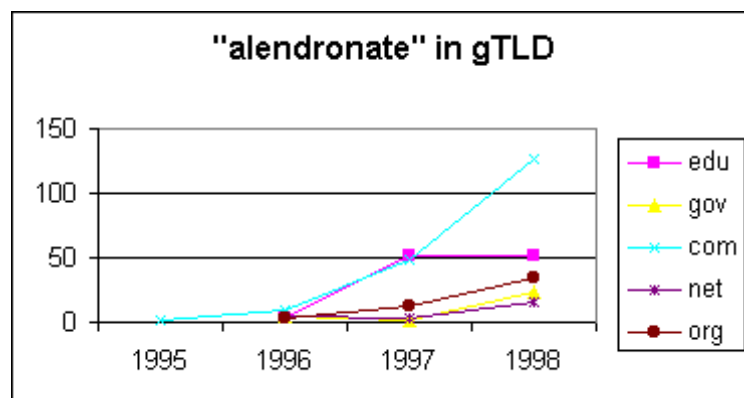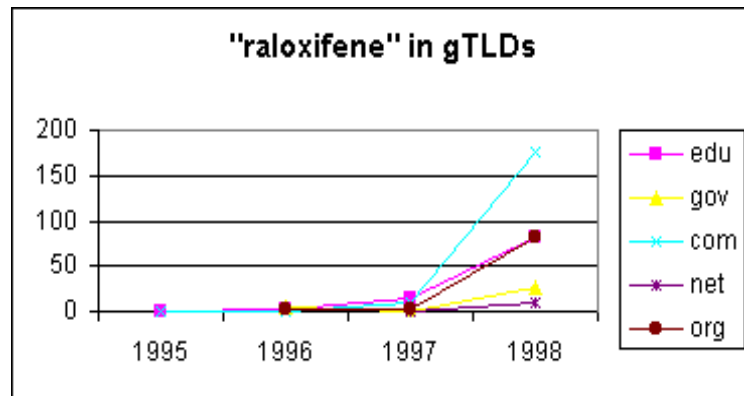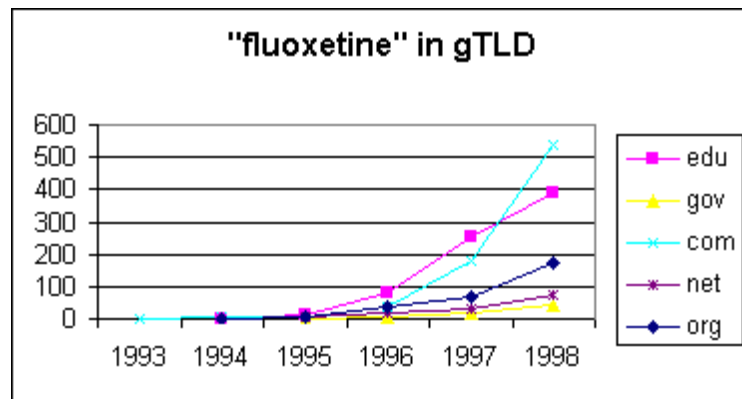
Figures 6a, 6b, 6c: Generic names of the three drugs at the Internet indicate knowledge-intensity

In conclusion, it seems that the Internet has become overwhelmed by the commercial and the marketing side of new developments, while the government sector ('.gov'), for example, has almost disappeared in relative terms **[vii]**. The '.edu'-domain, however, has remained relevant in size indicating a significant relation with the knowledge-production side when one uses generic names for the retrieval.

Note that although relatively insignificant, the number of hits in the case of the '.gov'-domain is still of an order of magnitude of half a million webpages. Thus, in absolute numbers these domains can remain substantive data sources. However, the number of webpages returned upon searching with, for example, 'domain:gov AND raloxifene' is only of the order of 25. While this result may be qualitatively informative, the representation is insufficient for drawing any conclusions about differences from other results. The value of indicators can in such cases be smaller than the error in the measurement. **[27]** For this reason, the rapidly growing Internet as such may be declining in value for the study and the comparison of various aspects of innovations. The 1990s has perhaps provided us with a specific opportunity for searching the Internet on this type of interfaces.

Can these hits be considered as indicating also user groups? We explored a few indicators here. First, one can combine the commercial trade names (e.g., Evista) with the non-commercial domains '.org' and '.net.' These searches led to relatively low numbers of hits as above indicated in the figures 6a and 6b. Another strategy is to limit the search to national domains and/or national languages, since patient organizations can be expected to emerge nationally **[13]**,**[26]**. However, we found no single hit for Evista nor Fosamax in the Dutch case (until and including 1998), and only 25 for the domain of Brazil and the case of Evista. When searching with Prozac as a keyword the number of hits remained also under one hundred per year in both these domains. It seems that the user is no longer playing an visible role at the Internet other than as a potential client **[viii]**. The Internet is nowadays highly commercialized.

7. The European Patent Office database

A next step in the reverse communication from the market side to the knowledge base of the innovations under study can be provided by the existing patent databases using the generic names. We used the database of the European Patent Office, since this website provides direct access to the complete file of World Patents including more than 30 million patents on-line (at **http://ep.espace.com**). Using this database, the following table of number of hits could be composed using the different keywords:

Keyword	Title words	Including also abstracts
Evista	1	1
raloxifene	42	51
Foxamax	0	0
alendronate	88	104
Prozac	3	5
fluoxetine	97	138

Table 2. Numbers of hits in the retrieval using trade names and generic names for the three drugs under study, in the fields of title words and abstract words, respectively.

Obviously, the generic words are used and the trade names are virtually non-existent in this representation[xi]. The single occurrence of Evista in the database relates to a 1999 patent for a Clematis plant with this same name [x] .The few patents using 'Prozac' in their title or as an abstract word refer to applications in which fluoxetine is used in combination with other drugs or in the context of a specific treatment. In these cases the word 'Prozac' is sometimes added between brackets behind 'fluoxetine.'

In summary, patents are organized within the database using exclusively the generic names of drugs. The incidents which use trade names are of an applicational nature and can be disregarded for the exploration of this knowledge base.
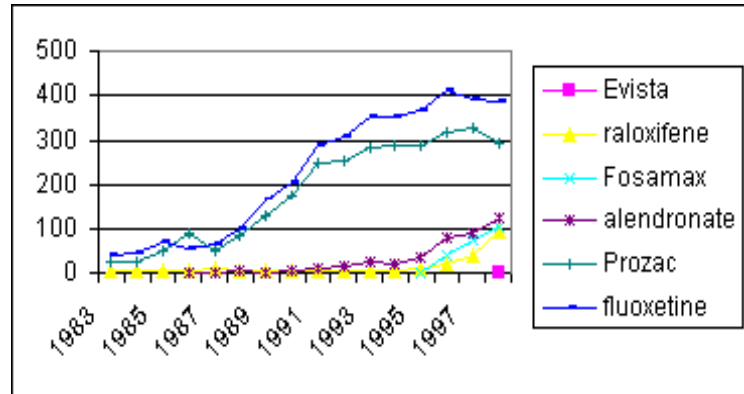
8. Medline

A database with relevance both in research and in medical practices is provided by Medline. Medline is nowadays fully available at the Internet at **http://www.ncbi.nlm.nih.gov/PubMed**. In this database, the trade names are always used in combination with the corresponding generic names, but the latter names also retrieve documents with medical applications other than the respective drugs themselves.

'Raloxifene', for example, is present in this database from the year (1983) that it was patented, that is, more than a decade before the invention of the trade name. 'Alendronate,' the active substance in Fosamax can be retrieved from 1986 onwards, while the first patents with 'alendronate' in the title are only from 1993. In other words, the Medline database exhibits the dynamics of the two competing compounds—with different company properties and university-industry relations—in their early stages, while the patent database does not.
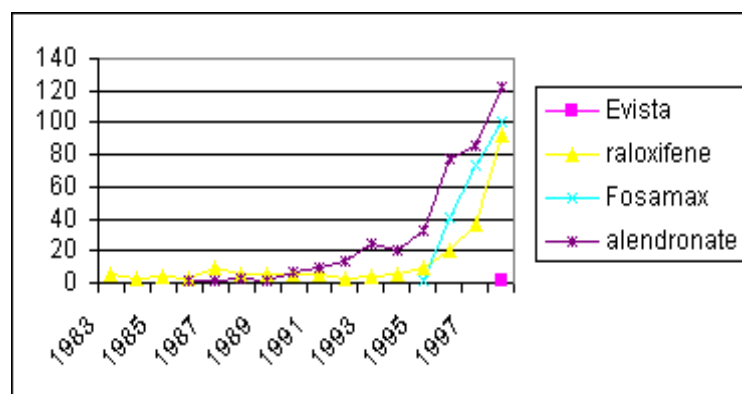
Figure 7a. Presence of Fosamax and Evista in Medline

Figure 7b. Presence of Prozac and fluoxetine in Medline

Remarkably enough, both the commercial name Prozac and the active chemical 'fluoxetine' which it contains, have been present as search terms in the Medline database from the 1970s onwards, while the patent was only published in 1988. Figure 7b exhibits these trendlines.

9. Word frequency distributions

Before moving onwards to the respective citation patterns, let us first make a comparison in terms of the vocabularies used in the three databases which we have discussed hitherto: the commercial side as retrieved by using keywords with the AltaVista search engine at the Internet, versus using the generic names in Medline and the World Patent database. For this comparison, I used for this comparison the data for the year 1998.

At the date of the comparison (7 May 2000), the search term Evista provided 273 hits using the Advanced Search Engine of AltaVista against 334 when searching with the generic name 'raloxifene.'[xi] I used the title words of the 228 pages in English among the 273 that could thus be retrieved globally. Among them eight pages could not be downloaded; the remaining 220 titles contained 387 unique words, of which only 185 occurred more than once. Fourteen titles were completely unrelated to any other in this set in terms of co-occurrences of title words. Actually, 51 words are meaningful to 154 of the cases. The ten or so most frequently used words are listed in the left-hand column of Table 3.

The right-hand column of Table 3 provides a similar listing for the 92 documents retrieved from Medline in this same year (1998) using 'raloxifene' as the search term. The semantic difference between the two lists corresponds to one's intuitive understanding of the functions of these different interfaces: the Medline words indicate the interests of the medical profession, while the AltaVista set informs us in accordance with the expected interests of potential users of the drug. The relative frequency distributions in the overlap between the two complete word lists, however, are significantly correlated (at the 0.001-level).[xii] Thus, the demarcation in terms of words used cannot be considered as statistically significant in this case.

f	Evista at the Internet	f	Raloxifene in Medline
27	New	34	Raloxifene
25	Drug	28	Estrogen
23	Line	24	Receptor

21	Medicine	19	New
20	Osteoporosis	15	Selective
15	Cancer	14	Modulator
14	Health	14	Osteoporosis
11	Breast	13	Effect
9	Raloxifene	11	Prevention
8	Estrogen	10	Tamoxifen
8	Women	8	Post-menopausal
7	Trial	8	Women
7	Update	7	Breast
		7	Cancer

Table 3. Words occurring most frequently at the Internet and in Medline

When we include into this comparison of title words also the much smaller set of 36 unique words contained in the five patents granted in 1998 with 'raloxifene' in the title, we retain an intersection of only thirteen words, of which seven are prepositions and articles.

Correlations:	ALTAVISTA	MEDLINE	EPO
ALTAVISTA	1.0000	.6973*	.5142
MEDLINE	.6973*	1.0000	.8116**
EPO	.5142	.8116**	1.0000
N of cases: 13	1-tailed Signif:*	- .01 **	- .001
without the seven articles and prepositions:			
Correlations:	ALTAVISTA	MEDLINE	EPO
ALTAVISTA	1.0000	.3372	.5457
MEDLINE	.3372	1.0000	.9638**
EPO	.5457	.9638**	1.0000
N of cases: 6	1-tailed Signif: *	- .01 **	- .001

Table 4: Pearson correlations among word lists from patents, medical files, and the Internet data (with and without articles and prepositions).

Table 4 provides the Pearson correlations between these frequency distributions. This table illustrates the well-known effect of common words (like prepositions and articles) generating correlations among otherwise different sets[30]. When these common words are excluded from the analysis, the correlation between the patent and the Medline databases are enhanced as an indicator of their common reference to the knowledge base. The correlation between the words from the Medline and the patent data, on the one side, and the Internet searches, on the other, is no longer significant under this condition (lower part of Table 4).

In summary, the linguistic variation enables the communication to reach out to new audiences, while the knowledge base also develops its restricted vocabulary. The coupling through patents is 'thin' at each moment in time: patents are rare events. Thus, the two communication circuits (that is, the market and the knowledge base) use another axis for the codification or, in other words, use a decomposable subdynamics of communication.[21]

The codification in these recursive systems of communication, however,

cannot be expected to function only linguistically (that is, in terms of the variation). Language users select on the basis of specific meanings and cognitions along the respective axes; for example, in terms of citations that enable scientific communicators to orient themselves among the complexities of agencies and communications.4 I will focus on this dimension of the codification in the remainder of this study. The discussion of codes and potentially small worlds at the Internet entails, for example, the use of tags and meta-tags, but this elaboration would lead me away from my research question about retrieving the science-base of innovation systems.

Can the patents perhaps be considered as punctuated equilibria between these communication circuits? Are these specific events and recombinations the sources of innovation (Langford, personal communication)? In my opinion, one should keep in mind that selection pressures always prevail. A single occurrence (like in a punctuated equilibrium) cannot be expected to suffice for the longer-term survival of an innovation:[31] the patents can therefore be expected to contain also an internal axis of recursive codification. I will now compare science and patent citations to explore the differences between the different types of codification from the evolutionary perspective of leading to these new drugs as innovations.

10. Two Citation Indices

Both the scientific and the patent literature use citations for the indication of intellectual lineages. References enable authors to shortcut elaborate discourse.[32] The codified indices pack the database so that storage, retrieval, and recall can be made more efficient. However, in a coded system knowledgeability and skills are needed for the recognition. Thus, specific competencies can be historically delineated into communities of professionals. Furthermore, the interactions between codes in scientific and patent literatures (e.g., cross-references) are not expected to be symmetrical since different codes can be implied in the selection on either side.[xiii],[33],[34]

While scientific citation is left to the discretion of the scientific author, the citation on the cover page of a patent is attributed by the patent examiner. The latter can build on the citations provided by an applicant in the full text of the application. However, the examiner has the obligation to check whether the claim is original by positioning the paper with reference to 'the state of the art.' Citations in patent applications are very focused around patentability ('prior art'). Thus, the citations of previous patents and non-patent literature provide us with indicators of highly focused selection routines by both applicants and examiners.[35]

a. The Science Citation Index

Figures 8a and 8b indicate the presence of the same keywords as used above, but now in the domain of the Science Citation Index. Since I used the on-line version of this database, available at the Internet as the so-called Web of Science, the data reach back only to 1988. Figure 8a shows the presence of the generic names in titles at dates before the granting of the respective patents, more or less analogous to, but somewhat later in time than the results from the Medline database reported above.
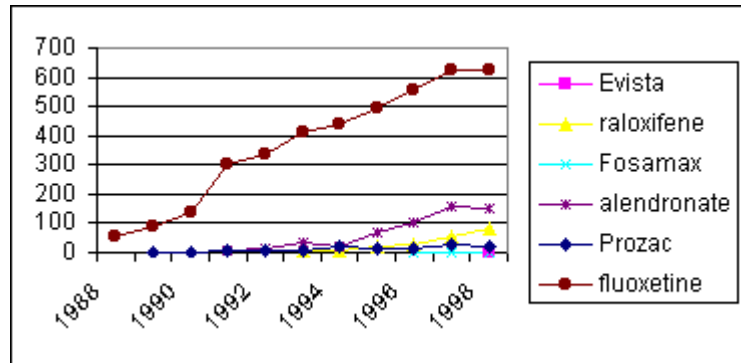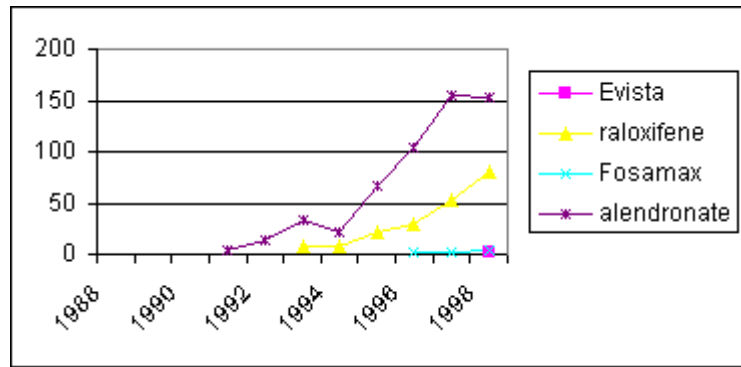
Figures 8a and 8b: Search results for the various trade and generic names in the Science Citation Index

Figure 8b shows that even after the dates of the respective patenting the trade names (e.g., Prozac) have not entered into the Science Citation Index to a significant extent. This is notably different from Medline. As a dedicated database, Medline maintains a window on the clinical side of the medical profession by using also the trade names. The Science Citation Index has remained exclusively research-oriented.

Note that 'alendronate' is represented in the SCI to an extent larger than 'raloxifene.' 'Alendronate' may have other applications which have not been shielded from the public arena by patent protection to an extent like that of Evista.

b. The Derwent Patent Citation Index

The search results from the Patent Citation Index are summarized in the following Table 5:[xiv]

|	raloxifene	Evista	alendronate	Fosamax	fluoxetine	Prozac
dates of original patents		1997		1995		1988
Number of patents	5		14		41	
authors	105		34		81	
nr of patent equivalents	550		108		251	
cited patents	604		150		495	
references	1237		183		450	
citing patents	120		19		168	

Table 5. Patent citation searches using the generic names of the three

compounds under study

Table 5 first indicates a pattern of heavy patenting on the side of Eli Lilly when compared with MSD for the case of alendronate (Fosamax). The middle column is considerably lower on all the parameters indicated. (Yet, a single patent may be commercially more important than a whole set.)

I shall now focus on the dynamics of the citations as an instrument to backtrack into the respective knowledge bases of these patents. As noted above, the references to patents within the scientific literature were relatively insignificant as compared to the extensive citation of both previous patents and journal articles within the patent database. Furthermore, one could observe a tendency to cite the original patent. However, the mechanism of codification is completely different between these two literatures.

To what extent do the patents refer to previous patents and to scientific literature? Figure 8 exhibits the age of the patent citations[xv]. First note that patents with reference to the two drugs marketed by Eli Lilly are more deeply rooted in previous patents than patents retrieved with the search term 'alendronate' (that is, Fosamax).
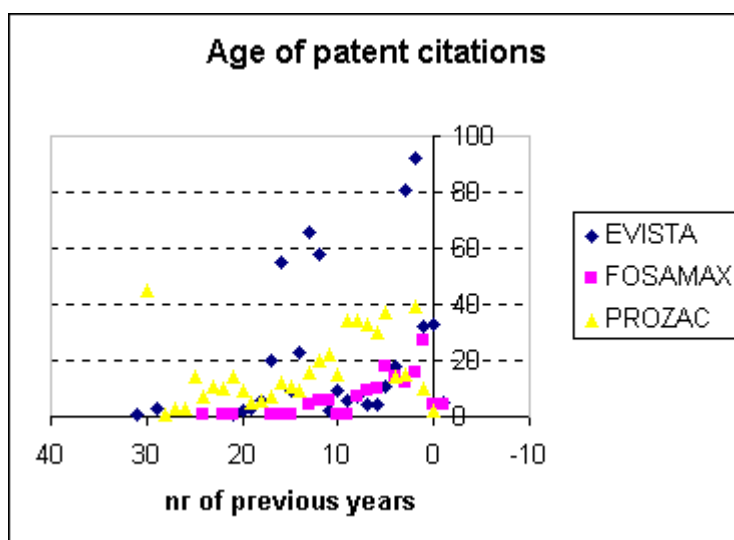
Figure 9. Age of patent-to-patent references in the case of the three compounds / drugs

Figures 10a and 10b provide the analogous figures for the scientific citations within patents. Since these so-called 'non-patent literature citations' are not completely standardized, the figures are based on a computer routine and therefore statistical: the scientific references within the patents were assessed for the cases in which the year was indicated either by ', 19??' or by '(19??)'. In practice, however, this routine covers almost all the data.
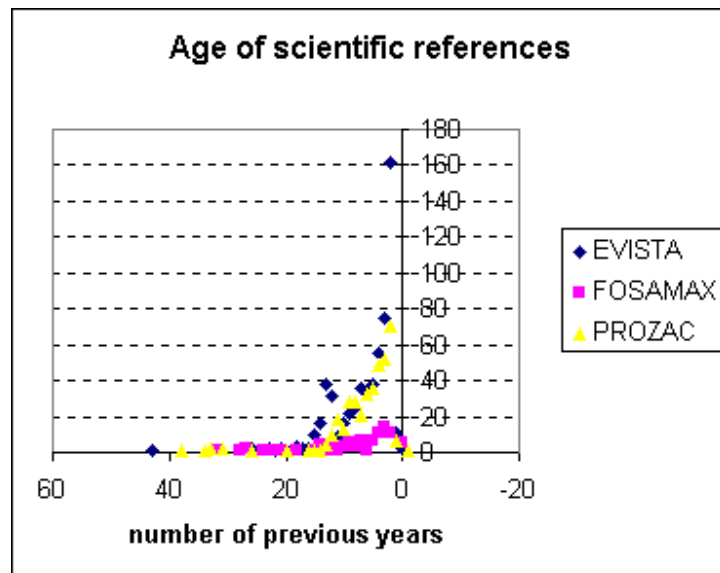
Figure 10a. Age of scientific references for patents referring to the three compounds
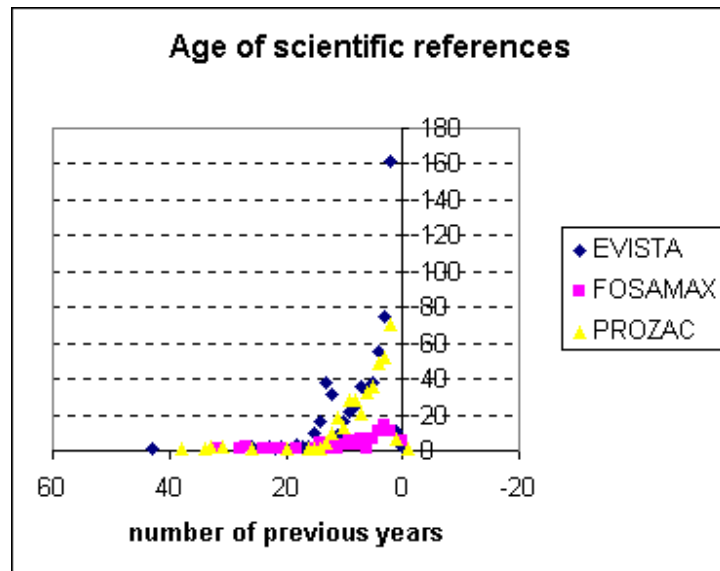
Figure 10b. Age of scientific references for patents referring to the three compounds, using a logarithmic scale.

Figure 10b provides a logarithmic representation of the curves in figure 10a. This enables us to see more clearly the differences in the slope between the two patents of Eli Lilly and the one of MSD in this case. As expected, all three lines show that the patent literature draws mainly on the *short-term memory of the research front*[36]. Citations of older literature are rare, much rarer than citations of older patents (see Figure 9). These results suggest that the recursive axis of the knowledge production system among patents is more important than interaction with their respective science bases.

However, the patents related to 'alendronate' (*Fosamax*) seem to be less tightly coupled to the present than the patents in the domains of the drugs produced by Eli Lilly. The 'alendronate' case of MSD is involved in the scientific knowledge base to a larger extent than the other two compounds that have been so important for Eli Lilly. This confirms the impression above that the latter company has been able to shield its core competencies from the academic community and university-industry relations more effectively than

MSD under otherwise comparable conditions of competition.

11. Conclusions

Our objective in this study has been to investigate whether trade names versus generic names of drugs can be used as indicators of 'Mode 1' and 'Mode 2' communication in the production of scientific knowledge and its application in knowledge-based innovations. While codification within the scientific knowledge base is known to provide indices in the communication both in terms of co-words and co-citations, the nature of the ongoing codification at the Internet is less obvious.

From an evolutionary perspective, codification is a necessary process in communication systems:[3] variation cannot provide all possible combinations, and existing channels of communication will increasingly shape pathways. The path-dependency leads necessarily to lock-ins,[21] to trajectory and niche formation,[31] potentially followed by globalization and regime (or paradigm) formation[37] .

The complex system of communications is composed historically by recombining different subdynamics, but evolutionarily it tends to be reshaped into functional axes under prevailing selection pressures. I distinguished above (i) the recursive axis of the (historical) production of new scientific knowledge, from (ii) the interface with the market in diffusion processes at each moment in time, and from (iii) the reflexive function of control both in the private (managerial) sphere and by public agencies (e.g., the FDA).

One unexpected, yet important conclusion has been that the Internet is nowadays so overwhelmingly commercial that it seems no longer useful as an indicator of 'user' interests. Both patient organizations and public health authorities have become marginal in terms of the representation. Thus, the current issue of 'social accountability' in innovation policies can no longer be covered adequately by relying on the Internet. Although patenting is obviously a regulatory function of the state,[12],[38] the dynamics of this public function of the state ('.gov') can only marginally be retrieved using the trade or the generic names of the drugs as search terms.

In the patent and science databases, the generic names prevail, with the exception of Medline which entertains an intensive relation with medical professionals and therefore adds the drug names into the searchable fields whenever applicable. The trade names could be retrieved at dates before the patenting. The knowledge-based innovations were thus visible in this (dedicated) database at the earliest moment in time. Patents seem to be a late indicator, but one can probably reconstruct the historical developments only with hindsight, that is, after that the patent has been granted, since the previous uncertainty (contained in the variation) can then selectively be provided with meaning for the perspective of the innovation as a result.

The internal dynamics of patent literature differs from that of scientific literature, although citation indices provide coupling mechanisms. This coupling is asymmetrical: patents seem to draw mainly on the current research front, while scientific literature seems to show a preference for citing the fundamental patents underlying the current applications at the commercial end. In other words, the scientific literature uses patents differently from scientific citations and patents use scientific citations differently from patent citations.

The various codes can also be considered as language variants with different functional—as opposed to national or regional—dialects. Scientific literature uses a language coded differently from that of patenting. The translation processes among different languages are further reflected by the interface between generic versus trade names at the interface with end-users. The

translation tends to black-box the internal dynamics of the knowledge production process in accordance with the competitive aims of the corporation in question. Throughout this study, however, we have been able to note also important differences between Eli Lilly and MSD as carriers of their innovation networks.

Acknowledgements

I wish to acknowledge valuable suggestions for this research by Mary Ellen Mogee, Marta Riba-Vilanova, and Cooper Langford.

Notes:

[i] The U.S. Patent and Trade Office makes the citation data available in ASCII format at **http://www.uspto.gov/patft/index.html** . However, the extraction of the relevant citations is not *sine cure*.[5, 6, 7, 8]

[ii] Although the Internet has a longer history, browsers (like Mosaic and then Netscape) have only been available since 1992. Before 1993 most files contained only plain text, and one is not able to retrieve hypertext structures from these texts with hindsight.

[iii] I used the expanded version of this database on-line at ISI's so-called Web of Science at **http://wos.library.tudelft.nl**

[iv] As against Evista, it is claimed that alendronate (Fosamax) not only prevents osteoporesis, but also reverses the process of decalcification of bones.

[v] See at **http://doc.altavista.com/adv_search/syntax.shtml** for the search syntax. The Powersearch engine of Northern Light has comparable search abilities, but they are less clearly organized.

[vi] Prozac has also become part of the common language. A university web server, for example, stated that "No one who uses the machine takes Prozac (as far as I know) or knows anything about it."

[vii] The '.gov' domain specifically refers to U.S. government agencies, while some of the other generic domains (e.g., '.com') are to some extent international.

[viii] Patient oriented websites can, however, be found embedded as individual pages at commercial portals like 'geocities.com'. the latter contains, for example, a category of websites for 'health and wellness' support.

[ix] These numbers include similar patents in different systems. The number of patents in Table 5 (below; that is, using the Derwent Patent Citation Index) is corrected for the double-counting.

[x] This is a European invention, but the patent was also applied for in the U.S.A. (US10932P).

[xi] Note that these values were 351 and 483, respectively, when the same searches were performed in November 1999. These differences can both be reflections of changes in the database, e.g., updates of websites that were previously dated as 1998 pages,[27] and general changes in the quality of the representation by the search engine over time.[26]

[xii] The Pearson correlation is .5686 in case that one uses the 61 words shared among the two sets; elimination of 16 prepositions and articles leads

to a Pearson correlation of .4666 which is still highly significant. Rank-order correlations were also significant.

[xiii] The asymmetry can be considered as analogous to the breach of symmetry by citations and references in the recursive dynamics along the time axis.[32, 33]

[xiv] The Derwent Patent Citation Index includes both examiner and applicant citations, but they can be searched also independently. In this study we did not further distinguish between the two types of citations. For a systematic comparison between these two types of citation, see, e.g., Meyer.[34]

[xv] In this stage, all citations in the patent, that is, both the examiner's and the applicant's ones, were included.[34]

References:

[1] Gibbons, Michael; Limoges, Camille; Nowotny, Helga; Schwartzman, Simon; Scott, Peter & Trow, Martin (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*, Sage, London, 1994.

[2] Cowan, Robin & Foray, Dominique (1997). The Economics of Codification and the Diffusion of Knowledge. **Industrial and Corporate Change**, 6:595-622.

[3] Leydesdorff, Loet (2001). A Sociological Theory of Communication: *The Self-Organization of the Knowledge-Based Society*. Universal Publishers <**http://www.upublish.com/books/leydesdorff.htm**>.

[4] Leydesdorff, Loet (1998). Theories of Citation?. **Scientometrics**, 43:5-25.

[5] Narin, Francis, & Olivastro, David (1988). Technology Indicators Based on Patents and Patent Citations. In: A.F.J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*, Elsevier, Amsterdam, pp. 465-507.

[6] Narin, Francis, & Olivastro, David (1992). Status report: linkages between technology and science, **Research Policy**, 14:237-249.

[7] Grupp, Hariolf (1996). Spillover effects and the science base of innovations reconsidered: an empirical approach. **Journal of Evolutionary Economics**, 6:175-197.

[8]Grupp, Hariolf, & Uwe Schmoch (1999). Patent statistics in the age of globalisation: new legal procedures, new analytical methods, new economic interpretation. **Research Policy**, 28:377-396

[9] Mogee, Mary Ellen, & Kolar, Richard G. (1999). Patent co-citation analysis of Eli Lilly & Co. patents. **Exp. Opin. Ther. Patents**, 9 (3):291-305.

[10] Nelson, Richard R. (1994). Economic Growth via the Coevolution of Technology and Institutions. In: Loet Leydesdorff and Peter Van den Besselaar (Editors), *Evolutionary Economics and Chaos Theory: New directions in technology studies*. Pinter, London, 1994, pp. 21-32.

[11] McKelvey, Maureen D. (1996). Evolutionary Innovations: *The Business of Biotechnology*. Oxford University Press, Oxford.

[12] Webster, Andrew & Rappert, B. (1997). Regimes of ordering: the

commercialization of intellecutal property in industrial-academic collaborations. **Technology Analysis and Strategic Management**, 9:115-129.

[13] Leydesdorff, Loet & Curran, Michael (2000). Mapping University-Industry-Government Relations on the Internet: An Exploration of Indicators for a Knowledge-Based Economy. **Cybermetrics**, 4 (2000), Issue 1, Paper 2 at **http://www.cindoc.csic.es/cybermetrics/v4i1p2.htm**

[14] Blauwhof, Gertrud (1995). *The non-linear dynamics of technological developments: an exploration of telecommunications technology*. Ph.D. Thesis, University of Amsterdam, 1995

[15] Meyer, Martin (2000). S&T indicators trapped in the Triple Helix? The case of patent citation citations in a novel field of technology. Paper presented at the *Third Triple Helix Conference on university-industry-government relations*, Rio de Janeiro, April 2000

[16] Zelman, Andrés & Loet Leydesdorff (2000). Threaded Email Messages in Self-Organization and Science & Technology Studies Oriented Mailing Lists. **Scientometrics**, 48: 361-380

[17] Lundvall, Bengt-Åke (1998). Innovation as an interactive process: from user-producer interaction to the national system of innovation. In: G. Dosi, C. Freeman, R.R. Nelson, G. Silverberg and L. Soete (Eds.), *Technical Change and Economic Theory*. Pinter, London, 1988, pp. 349-369.

[18] Andersen, Esben Slot (1994). *Evolutionary Economics: Post-Schumpeterian Contributions*. Pinter, London, 1994.

[19] Leydesdorff, Loet & Van den Besselaar, Peter , Eds. (1994). *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies*. Pinter, London, 1994

[20] Etzkowitz, Henry & Leydesdorff, Loet (2000). The Dynamics of Innovation: From National Systems and 'Mode 2' to a Triple Helix of University-Industry-Government Relations. **Research Policy**, 29:109-123.

[21] Arthur, W. Brian (1989) Competing Technologies, Increasing Returns, and Lock-In by Historical Events. **Economic Journal**, 16:116-131.

[22] Blume, Stuart S. & Geesink, Ingrid (2000). Vaccinology: a science and its problems. **Science as Culture**, 9: 41-72

[23] Simon, Herbert A.(1973). The Organization of Complex Systems. In Pattee, H. H. (Ed.), *Hierarchy Theory: The Challenge of Complex Systems*, George Braziller Inc., New York, 1973, pp. 1-27.

[24] Luhmann, Niklas (1984). *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Suhrkamp, Frankfurt a. M., 1984. [*Social Systems*, Stanford University Press, Stanford, 1995.]

[25] Mogee, Mary Ellen (1999). Mapping Technology with Patent Databases. Paper presented at the *Annual Meeting of the Society for Social Studies of Science*, San Diego, 30 October 1999

[26] Boudourides, Moses A.; Sigrist, Beatrice , & Alevizos, Philippos D. (1999). Webometrics and the Self-Organization of the European Information Society, at <**http://hyperion.math.upatras.gr/webometrics**> (26 October 1999).

[27] Rousseau, Ronald (1999). Daily time series of common single word searches in AltaVista and NorthernLight, **Cybermetrics**, 2/3 (1999), Paper 2. <**http://www.cindoc.csic.es/cybermetrics/v2i1p2.htm**>.

[28] Butler, Declan (2000). Souped-up search engines, **Nature**, 405:112-115

[29] Leydesdorff, Loet (2000). A Triple Helix of University-Industry-Government Relations. **The Journal of Science & Health Policy**, 1: 43-48.

[30] Salton, G. & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill, Auckland, etc.

[31] Bruckner, Eberhard; Ebeling, Werner; Jiménez Montaño, Miguel A..& Scharnhorst, Andrea (1994). Hyperselection and Innovation Described by a Stochastic Model of Technological Evolution. In: Loet Leydesdorff and Peter Van den Besselaar (Eds.), *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies*. Pinter, London, 1994, pp. 79-90.

[32] Bernstein, Basil (1971). Class, Codes and Control, Vol. 1: *Theoretical studies in the sociology of language*. Routledge & Kegan Paul, London, 1971.

[33] Fujigaki, Yuko (1998). Filling the Gap Between Discussions on Science and Scientists' Everyday Activities: Applying the Autopoiesis System Theory to Scientific Knowledge, **Social Science Information**, 37:5-22.

[34] Fujikagi, Yuko & Leydesdorff, Loet (2000). Quality Control and Validation Boundaries in a Triple Helix of University-Industry-Government Relations: 'Mode 2' and the Future of University Research. **Social Science Information**, 39 (4):635-655

[35] Meyer, Martin (2000). What is special about patent citations? Differences between scientific and patent citations. **Scientometrics**, 49:93-123

[36] Price, Derek J. de Solla (1965). Networks of scientific papers. **Science**, 149:510-515.

[37] Leydesdorff, Loet & Van den Besselaar, Peter (1998). Technological Development and Factor Substitution in a Non-linear Model. **Journal of Social and Evolutionary Systems**, 21:173-192

[38] Van den Belt, Henk, & Arie Rip, The Nelson-Winter (1987). Dosi model and synthetic dye chemistry. In: Wiebe Bijker, T. P. Hughes, and T. Pinch (Eds.), *The Social Construction of Technological Systems*, MIT Press, Cambridge, MA, 1987, pp. 135-58