

# **In search of anti-commons: patent-paper pairs in biotechnology. An analysis of citation flows.**

**Tom Magerman<sup>a,b,c</sup>**

**Bart Van Looy<sup>a,b,c</sup>**

**Koenraad Debackere<sup>a,b,c</sup>**

<sup>a</sup> *Managerial Economics, Strategy and Innovation, Faculty of Business and Economics,  
K.U.Leuven, Leuven (Belgium)*

<sup>b</sup> *Research Division INCENTIM, K.U.Leuven, Leuven (Belgium)*

<sup>c</sup> *ECOOM, Leuven (Belgium)*

**Extended abstract**

**Keywords:** science-technology linkages; entrepreneurial university; patent-paper pairs;  
anticommons

## **Introduction: Entrepreneurial Universities**

While most empirical evidence – on the level of individual scientists – reports a positive relationship between patenting activities and publication outcomes (quantity as well as quality), the expansion of IPR might still result in ‘privatizing’ the scientific commons and potentially limiting scientific progress (Argyres and Liebeskind, 1998; David, 2000; Krimsky, 2004). This fear is nicely expressed by the metaphor of the “Tragedy the anti-commons”, introduced by Heller (Heller, 1998) as opposed to the “Tragedy of the commons” of Hardin (Hardin, 1968). Heller states that the presence of too many owners with blocking power can lead to the underutilization of scarce resources, or, translated to the world of IPR, more intellectual property rights may lead paradoxically to fewer useful products (too many owners hold rights in previous discoveries creating obstacles for future research).

Although anecdotal evidence exists of problematic use of IPR on scientific findings (e.g. the ‘OncoMouse’ or ‘Harvard mouse’ of Leder and Stewart; patents on human genes associated with breast and ovarian cancer owned by Myriad Genetics), large scale evidence of the presence of an anti-commons effect in biotechnology patenting is rare. One notable exception is the study of Murray and Stern (2007) suggesting a modest anti-commons effect based on a decline in citation rate – after granting of the patent - by 10 to 20% for a set of 169 patent-paper pairs published in Nature Biotech between 1997 and 1999, although these authors also clearly point to the interpretation limits inherent to their study.

In our study, we want to contribute to the research on an anti-commons effect in biotechnology by comparing citation patterns of patents and scientific publications for a large dataset containing all biotechnology patents (EPO and USPTO) and scientific publications (published in ISI Web of Science covered journals) from 1991 to 2008. First we investigate whether biotechnology publications for which a counterpart exists in the patent system (so called ‘patent-paper pairs’, scientific publications from which the contents - methodology, findings, discovery - is part of a patent application) are cited differently (more/less) within scientific journals, compared to similar biotechnology publications which are not related to a patent document. Next, we engage in a similar analysis focusing this time on ‘technological’ citations: to what extent are patents closely related to scientific publications cited differently by other patents compared to biotechnology patents without scientific counterpart. The former will allow us to shed some light on the fear that exploitation of scientific findings is hampering scientific development by pruning promising developments due to the

introduction of (potentially blocking) patents. The latter will allow us to look at the technological impact of scientific developments that become translated into a patent.

## **Data and methodology**

On the patent side, the OECD definition of biotechnology is used to identify biotechnology patents (OECD, 2005). We use PATSTAT (EPO Worldwide Patent Statistical Database) to retrieve all EPO and USPTO granted patents with application and grant year between 1991 and 2008 according to the 30 defined IPC-subclasses/groups related to biotechnology (88,248 patents in total).

On the publication side, we select biotechnology publications (articles, letters, notes, reviews)<sup>1</sup> from the Thomson Reuters ISI Web Of Science database based on the Web of Science subject classification, for the same time period 1991-2008 (948,432 biotechnology related publications).

An important methodological aspect for this kind of studies relates to the identification of those patent-paper pairs, scientific publications for which a patent equivalent is present. To obtain a broad set of patent-paper pairs, we stepped down from a manually guided process of mapping patent and scientific publications and developed a new approach of automated, large scale, mapping of patents and scientific publications based on content similarity by relying on text mining algorithms (Latent Semantic Analysis - Landauer et al., 2007).

For all patents, the similarity with all publications is derived based on 43 content similarity metrics. A thorough manual validation of 300 cases was set up to select the metric best suited for the identification of patent-paper pairs. This validation effort revealed that dimensionality reduction as advocated by methods like Latent Semantic Analysis underperforms compared to a normal cosine measure on the full data, and measures based on a mere count of the number of common terms yields the best and most robust results in terms of identifying patent-paper pairs without missing relevant pairs.

Two content-based criteria are combined with an additional criterion based on ownership to identify 584 patent-paper pairs.

---

<sup>1</sup> Articles are by far the biggest category (90% articles compared to 1.5% letters, 2% notes and 6.5% reviews)

## Findings

### *Citation patterns of scientific publications (publication-to-publication citations)*

On average we clearly observe substantially higher forward citation counts for publications that are part of a patent-paper pair and other publications (mean of 161.8 versus 46.9, median of 65 versus 20).

To verify the significance of the observed difference when controlling for other factors multivariate analysis have been performed. Given the nature of the data (citation data) we opted for a negative binomial regression with the number of forward citations as dependent variable and a dummy variable indicating whether a publication is part of a patent-paper pair as independent variable.

To adjust for the expected difference in average quality between paired and non-paired publications (due to the potential selection bias of publications that are part of a patent-paper pair), we only include publications from journals that have at least one publication that could be paired with a patent, i.e., we only use publications that are comparable in average impact factor because they originate from the same set of journals.

Regression results reveal that publications being part of patent-paper pairs have significantly more forward publication citations. One also notices a positive relationship between forward citations and the number of authors as well as the number of backward citations. The number of forward citations differ significantly between journals. Finally, the observed citation rates reflect an inverse U pattern over time. When removing outliers, i.e. all publications with a forward citation count larger than the mean plus three times the standard deviation, similar results are obtained.

Inspired by the observations of Murray and Stern (2007) - a relative decline in citation patterns after patents have been granted – we verify whether the citation rates differ before and after a patent has been granted.

The ratio of average citations received by pairs versus non-pairs equals to 1.71 and 1.74 before and after granting respectively. While these descriptive statistics do not indicate a decline, a formal t-test reveals that both ratios are not significantly different ( $p=0.86$ ). As such, our data do not show any sign of anti-common effects that become visible after patent rights have been granted.

### *Citation patterns of patents (patent-to-patent citations)*

In order to assess whether observed differences are statistically significant, we performed a negative binomial regression with the number of forward patent-to-patent citations as dependent variable and a dummy variable indicating whether a patent is or is not part of a patent-paper pair as independent variable.

Regression results reveal that patents being part of a patent-paper pairs have more forward publication citations, but the difference is not significant. USPTO patents have more citations than EPO patents. Other controlling variables have a significant and positive impact (presence of academic patentee; number of backward scientific non-patent citations; number of backward patent citations; number of forward publication citations from WOS publications; number of inventors), except for the number of patentees, which has a negative but not significant impact and time, which displays a decreasing, curvilinear relationship with patent citations.

After removing outliers, i.e. all patents with a forward citation count larger than the mean plus three times the standard deviation, similar results are. Finally, when we limit the time period to all patents applied for between 1991 and 2000 – in order to allow all patents to have at least 10 years of forward patent citations – patent-paper pairs have less forward patent citations, but also this difference is not significant (both when including and excluding outliers). Overall, we observe no significant difference in terms of (forward) patent citations when comparing patents that are associated with a scientific publication with their solitary counterparts.

### **Discussion and (intermediate) conclusions**

After studying the citation patterns of the documents included in the patent-paper pairs, we were not able to detect a significant anti-commons effect on the basis of the 584 pairs identified. These findings add to the current stock of insights on the interaction between patenting and publication behaviour. Through the design and application of advanced text mining techniques on a broad set of data, we intended to take the current insights a step further. Extensive validation efforts were undertaken in order to confirm the results obtained.

These results definitely are an invitation to further examine the joint effects of patenting and publishing activities by scientists. The first point of attention that arises is the one of generalization towards other fields of ‘techno-scientific’ economical activity. Can we

substantiate the current findings in technology domains such as materials or in other fields? The second point relates to corroborating and consolidating the robustness of the text mining methodology that was deployed, as well as a further, independent, confirmation of the optimal identification algorithm. The third point pertains to the continuous cross-validation of the results obtained with our method with the results obtained by sets of patent-paper pairs that have been constructed manually by experts.