RIETI Policy Discussion Paper Series 16-P-007

Does Standardization Affect Science Linkage? Implications for society and organizational management: Evidence from artificial intelligence-related technology

TAMURA Suguru
RIETI

IWAMI Shino
University of Tokyo

SAKATA Ichiro
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Policy Discussion Paper Series 16-P-007
April 2016
Revised: October 2017

Does Standardization Affect Science Linkage? Implications for society and organizational management: Evidence from artificial intelligence-related technology

TAMURA Suguru
Research Institute of Economy, Trade and Industry (RIETI)

IWAMI Shino
Policy Alternatives Research Institute (PARI), University of Tokyo

SAKATA Ichiro
Policy Alternatives Research Institute (PARI), University of Tokyo/RIETI

ABSTRACT

This study examines the influence of standardization in scientific technology research. For this purpose, we use the knowledge structure modeled by academic papers, patent filings, and standardization, and a clustering analysis for the research method. As the research subject, we studied image-digitizing technology de jure standard that is used to encode and decode the transmission of audio and video. The technology is widely analyzed by artificial intelligence technologies such as machine learning. Our study contributes in several ways. First, it provides a new knowledge structure consisting of (1) patents, (2) academic papers, and (3) standardization. Second, it determines whether (1) standardization has the influence to change the relationship between academic papers and patents in terms of a science linkage specifically relating to standardized technology and whether (2) standardization influences the academic productivity of researchers in the technology field.

***Keywords*:** Standardization, Science linkage, Watermark
JEL Code: C83, D83, L15, O34

This study is conducted as a part of research at the Research Institute of Economy, Trade and Industry (RIETI). The authors appreciate the valuable comments of Discussion Paper seminar participants at RIETI and support from Director Nagano, Director Fukuda, and Director Tokumasu at METI, support from Lecturer Junichiro Mori at the University of Tokyo and support of Innovation Policy Research Center at the University of Tokyo. This work was supported by JSPS KAKENHI Grant Number 15K03718.

1. INTRODUCTION

In this study, we examine the influence of an image-digitizing technology (MPEG[1]) that is increasingly analyzed using artificial intelligence-related technology. MPEG is one of the most influential standardized technologies in today's digital society. It is widely used for encoding and decoding audio and moving images. In this study, we analyze the influence of standardization on the science linkage of MPEG.

Historically, MPEG has been standardized and developed under section JTC1/SC29 of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) through the initiative of East Asian researchers led by Professor Emeritus Hiroshi Yasuda[2] of The University of Tokyo (formerly the director of the Nippon Telegraph and Telephone Corporation (NTT) R&D center) (Yasuda, 1989). We can observe the influence of standardization on these technologies because the ISO and the IEC have issued international standards in relation to these technologies (Yasuda, 1989). In addition, these technologies are widely analyzed in relation to image recognition by artificial intelligence technologies (Albusac, Castro-Schez, Lopez-Lopez, Vallejo, and Jimenez-Linares, 2009; Amato, Savino, and Magionami, 2008; Fernández, Kalva, Cuenca, and Orozco-Barbosa, 2007; Fernández-Escribano, Kalva, Cuenca, and Orozco-Barbosa, 2006). Such artificial intelligence-related technologies are now becoming a prime R&D target. They even include such technologies as *AlphaGo*, developed by Google DeepMind, which was the first computerized Go program to defeat a human player (Silver *et al*., 2016). In future, these technologies will result in social and organizational transformation (Colbert, Yee, and George, 2016; Felten, 2016). Accordingly, this transformation will become a key social and management issue. In summary, MPEG is a viable technology that warrants detailed examination from the social and management perspective.

We examine the exploration and exploitation of MPEG technology using bibliographic information. In this study, the concept of a vertical integration of exploration and exploitation strategies is explained. We examine a practical case and model of the vertical integration between exploration and exploitation. We also show how this strategy can enable new technological research and how organizations can use their newly discovered knowledge.

The balance between exploration and exploitation has been considered important by scholars since the early 1990s (March, 1991). Prior research has focused on the presumption of a horizontal integrated relationship between exploration and exploitation (He and Wong, 2004) rather than a vertical integrated relationship. This means that it is not necessarily empirically true in the current ICT environment that vertical exploration and exploitation strategies are theoretically more difficult than horizontal exploration or exploitation strategies. Prior research has focused on factors other than ICT (Enkel, Heil, Hengstler, and Wirth, 2016). This is the empirical research outcome of this study to the theoretical academic discussion. In summary, it is essential that scholars examine ways to implement the exploration and exploitation strategy rather than merely theorizing it.

To examine these issues, we address the following questions: "Does standardization affect science linkage?" and "What are the implications for society and organizational management?" For this purpose, using large data analysis with computation (i.e., bibliographic clustering of citation networks (BCCN)) for patents, academic papers, and standards relating to MPEG, we, to the best of our knowledge, first empirically examine the influence of standardization on science linkage. As a result, in a specific technology such as artificial intelligence-related imaging technology (i.e., MPEG),

---

[1] MPEG is the name of the standardized technology in the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) (Yasuda, 1989). MPEG is an abbreviation of Motion Picture Expert Group, which was the expert group involved in the preparation of the technology standard. Later, the expert group name was adopted as the name of the standardized technology. Hence, we use MPEG to mean the technology in this study. This technology is used to encode, decode, and subsequently convey video and audio files. Today, MPEG is widely used to deliver videos over the Internet.

[2] In 1996, he won an Engineering Emmy Award (US) from the Television Academy for his contributions.

standardization has an effect on science linkage and can change the knowledge flow between academic papers and patents. Further, this information can be used to exploit knowledge in pursuit of an R&D strategy. In summary, we identify 1) a promising R&D region (which corresponds to *exploration*), and 2) a way to use this R&D region (which corresponds to *exploitation* of newly obtained knowledge). We also analyze the results in an organizational management context.

The bibliographic clustering method is now widely used to study knowledge flow and the search for knowledge (Leydesdorff, 1995). Related research is accumulating as a result of the rapid development of ICT and bibliographic data infrastructure (Shibata, Kajikawa, Takeda, and Matsushima, 2008; Tashiro, Tashiro, Iwami, and Sakata, 2013). The main research subject is the relationship between patents and academic papers. Scholars call this knowledge flow "science linkage." Academic papers are the important result of scientific research, and are the source of new technology that is applicable to industry (Cohen, Nelson, and Walsh, 2002; Looy, Zimmermann, Veugelers, Verbeek, Mello, and Debackere, 2003), although scholars have not considered standardization to be an important factor in innovation for a long time now. Hence, they have not applied the bibliographic clustering method to the analysis of standardization, nor have they discussed the influence of standardization on knowledge accumulation. One reason for this is that there are no stylized data on standardization. Nevertheless, more recently, empirical analyses targeting the relationship among patents, research papers, and standards has begun to appear. According to recent studies, standardization has both positive and negative influences on the production of patents and academic papers in the basic research stage (Blind and Gauch, 2009; Zi and Blind, 2015). In this study, we aim to validate these findings.

Further, we use bibliographic analysis to observe organizational management issues. To our knowledge, this method is commonly used to examine science linkage itself, rather than to observe management issues. Giving consideration to organizational economics (Milgrom and Roberts, 1992), we use BCCN to analyze the multidimensional task problem (Holmstrom and Milgrom, 1991). Using this approach, we aim to connect bibliographic analysis with the relevant economic theories.

We find novel evidence of the influence of standardization on today's society and management. Our first contribution is to show the relevance of exploration *and* exploitation compared with exploration *or* exploitation using BCCN, which is dependent on unsupervised clustering. Further, we show the transition from a horizontal integration to a vertical integration between exploration and exploitation. Our second contribution is to provide a knowledge space structure with respect to standardization, in addition to patents and academic papers. Our third contribution is the discovery that, in relation to the specific technology related to MPEG, standardization can change the knowledge flow between academic papers and patents. For a long time now, it has been unclear how standardization activities influence R&D activities in research organizations (Tamura, 2012; Tassey, 2003; Zi and Blind, 2015). This implies that standardization can affect the productivity of researchers' academic publications in such technological domains. We capture issues of organizational economics such as the multidimensional task problem through this finding.

2. LITERATURE REVIEW AND HYPOTHESIS FORMATION

2.1 Overview

Bibliographic analysis of social and management issues is still emerging. Hence, in prior bibliographic research, scholars have not necessarily included all the essential elements for innovation. For example, the influence of standards has not been well considered. Shibata, Kajikawa, and Sakata (2010) made use of academic papers and patent filings for their bibliographic analysis. Nevertheless, they have not used standardization data. One reason for this is that there are still insufficient standardization data, even in this era of big data (George, Haas, and Pentland, 2014; Tamura, 2013). Another reason is that, in the existing corporate strategy and national innovation system, we regard 1) academic papers and 2) patents as the primary output indicators for R&D projects. Thus, there is no incentive to measure the influence of standardization. However, in recent national innovation systems in major research regions such as the US, the EU, and Japan, standardization has become increasingly significant in their innovation policy and corporate strategy. In the EU, it has now become an inevitable element in policy evaluation as well as R&D project evaluation (Edler, Georghiou, Blind, and Uyarra, 2012). In Japan, the government expects the policy regarding standardization to play an essential role (Cabinet Office, 2016). However, at present, it is not implemented sufficiently. In the US, a system to evaluate standardization remains insufficiently developed (Tassey, 2003). In Japan, with respect to the national project, the New Energy and Industrial Technology Development Organization has adapted the results of standardization, albeit only to measure the number of draft proposals. This means that in terms of management perspectives, it is still difficult to capture the level of standardization activities within research institutions and to reflect this in researchers' rewards.

2.2 Management Perspectives

2.2.1 Exploitation and exploration using bibliographic analysis

Bibliographic analysis can expand the ability of organizations to conduct exploration. It involves less uncertainty in terms of related costs, which have been identified as an obstacle to exploration (March, 1991). Further, it enables organizations to find a way to use the newly discovered knowledge. Sometimes, external knowledge is difficult for an organization to implement. This is another cause of uncertainty. Prior research has mainly focused on the horizontal integration between exploration and exploitation when considering the limited managerial resources in practice. A horizontal integration means that there should be a balance between exploration and exploitation in relation to organizational learning. However, the relationship between the two is not necessarily complementary in practice. The main reason for this is the uncertainty of exploration (March, 1991). In addition, it is difficult for an organization to evaluate the quality of discovered knowledge because of insufficient absorption capacity. Hence, they cannot determine how to use external knowledge, even when they find it. A combination of different types of knowledge is required for innovation (Alexander, 1964; von Hippel, 1994). For this purpose, firms search the knowledge space. This space has an internal sector (within firms) and an external sector (outside firms) (Grant, 1996; Mansfield, 1988; Rosenkopf and Nerkar, 2001). In addition to generating internal knowledge, firms can be innovative when they translate the knowledge around them to create new products (Katila, 2002). Therefore, knowledge located outside organizational boundaries plays an important role in firm performance (Laursen and Salter, 2006). Hence, we can apply BCCN to both internal and external knowledge searches.

Moreover, the breadth of the external search is positively related to individual innovativeness (Dahlander, O'Mahony, and Gann, 2016; Tortoriello, 2015). Nevertheless, exploration involves the risk of failure. Because of this, firms need to explore two different dimensions of organizational search, namely, breadth and depth (Katila and Ahuja, 2002; Laursen and Salter, 2006; March, 1991). These two dimensions can be seen as having a trade-off relationship. The relationship between exploration and exploitation was previously seen as a horizontal integration rather than a vertical integration. Thus, in the past, firms have depended on the use of similar technologies to produce new products (Helfat, 1994; Wade, 1996) and create an environment of path dependency in relation to innovation (Nelson and Winter, 1982). This can lead to a competency trap, whereby exploration is avoided (Levitt and March, 1988) and radical innovation is restricted. Further, to find the knowledge space outside a firm's boundary is a difficult task. Organizations are not cognitive of the entire knowledge space. Therefore, the knowledge they find is sometimes incomplete and less than they require, even though they may think it is complete (Kahneman and Tversky, 1979). Hence, their decisions are *bounded rational* (Simon, 1979).

Further, timely detection of new technological frontiers leads to the attainment of a first-mover advantage in terms of R&D strategy. In addition, firms usually do not create disruptive innovation intentionally. Hence, new concepts are recognized and formed once the research has progressed to a certain stage in public. This means that finding ways to use these new concepts is also difficult. BCCN can help to find new technology in the knowledge space, as well as new ways to use it (Tashiro *et al.*, 2013).

In summary, a lack of information processing technology has, generally speaking, made it difficult for organizations to explore the knowledge space effectively and completely. Hence, exploration and exploitation have come to be seen as separate approaches, and organizations have traditionally chosen to pursue one or the other (March, 1991). Nevertheless, especially in terms of R&D strategy, the environment has changed as a result of the rapid expansion of data availability and the development of information processing technology (George *et al.,* 2014; George, Osinga, Lavie, and Scott, 2016). Using the bibliographic method that has been developed, the prior theoretical framework changes. Applying BCCN to the relevant documents reduces the cost of searching. Further, it is easy to visualize emerging concepts, which are not necessarily able to be described in existing terms. Therefore, we can discover new knowledge spaces arising from emerging technologies. This is in keeping with the idea that "Painting a 'big picture' of scientific knowledge has long been desirable…" (Borner, Chen, and Boyack, 2003). Now, exploitation and exploration can be compatible. Nevertheless, there are few studies examining this change in organizational ability, which has arisen from developments in ICT and the big data environment.

2.2.2 Multidimensional tasks in R&D

Generally, evaluating task achievement is a difficult process. For example, while quantity is easy to measure, creativity is difficult to measure (Kachelmeier, Reichert, and Williamson, 2008). Thus, in relation to employees, there is a multidimensional task problem (Holmstrom and Milgrom, 1991). This applies to the evaluation of academic achievement. The relationship between achievement and the key indicators is: Achievement = f (academic papers, patents, standardization), where $f'(\cdot)>0$. Of these factors, the number of papers published has a primary influence on researchers' job security, and we regard it as the most important factor in the reputation of researchers (Dewett and Denisi, 2004; Oyer, 2006). As for the relationships between factors, it was found that in the BAM Federal Institute for Materials Research and Testing in Germany, standardization activities are negatively correlated with journal publication in basic research, because preparation of academic papers is so labor intensive that researchers cannot spare any time for standardization activities (Zi and Blind, 2015). Conversely, scholars have reported that standardization helps to improve scientific research in the basic research stage (Blind and Gauch, 2009). This situation presents a dilemma for both researchers and management, because it relates to the career development of researchers. Under these conditions, are standardization activities compatible with the publication of academic papers? If not, there should be some form of support via a policy that encourages the improvement of standardization activities in academic institutions. Otherwise, this leads to a paradox whereby to improve basic research, we need standardization activities, but those standardization activities hinder academic achievement. Following prior results, we express the relationship between academic publication and standardization in public research institutions in Germany as:

Academic paper publication = – g (standardization), where $g'(\cdot) >0$.

In prior studies, researchers generally did not discuss the role of standards with respect to academic work (Funk and Luo, 2015). Hence, insufficient research has been undertaken on this topic. We address this issue in this study.

2.3 Methodological Perspectives

2.3.1 Science linkage between patents and academic papers

Patents are an important indicator of R&D success and innovation (Acs and Audresch, 1989; Griliches, 1984; Griliches, 1990; Jaffe, 1986; Salter and Martin, 2001; Trajtenberg, 1990a; Trajtenberg, 1990b). Further, the patent citation network contains information about patents and the links among them (Carpenter and Narin, 1983; Carpenter, Narin, and Woolf, 1981; Trajtenberg, Henderson, and Jaffe, 1997). Hence, this is an important source of data for bibliographic analysis. Similar to patents, citations among academic papers also provide important information. Garfield pioneered the use of citation analysis among academic papers (Garfield, 1955). Further, scholars study academic paper networks based on co-authorship to analyze knowledge flows (Demirkan, Deeds, and Demirkan, 2013). If patents are the private knowledge stream, academic papers are the public knowledge stream (Huang and Murray, 2009). Hence, the notion of science linkage examines the flow of knowledge from public knowledge to private knowledge. This is useful for predicting potential areas of technological development (Shibata *et al*., 2010). Thus, various studies have examined citation networks (Demirkan and Demirkan, 2012; McMillan, Narin, and Deeds, 2000; Oliver, 2004; Oliver and Liebeskind, 1997; Zucker, Darby, and Armstrong, 2002).

2.3.2 Knowledge space structure and standardization

In addition to the relationship between patents and academic papers, this study uses information on standardization, as shown in Figure 1. The knowledge space model includes 1) patents, 2) academic papers, and 3) standardization, in contrast to prior research, which only includes 1) patents and 2) academic papers (Shibata *et al*., 2010). Therefore, organizations can recognize standardization as a factor in the knowledge space. Further, using this model, we can observe how standards influence the relationship between the technological similarities of patents and academic papers. In previous studies relating to standardization, there has been little attention paid to bibliographic analysis (Arthur, 1989; David, 1985). However, standardization is now important, especially in fields such as ICT and the specific subject of this study; an artificial intelligence-related technology (Egyedi and Sherif, 2010; Jakobs, Procter, and Williams, 2001; Sherif, 2001). Thus, we include standardization as one of the essential components of the knowledge space in this study.

[Insert Figure 1 about here]

In terms of patents and standardization, prior research has found that standardization activities increase the number of patent applications (Zi and Blind, 2015). Nevertheless, in general, we do not patent standardized technology unconditionally (Tamura, 2016).

2.4 Hypotheses

From the above argument, we arrive at the following two hypotheses:

Hypothesis 1 (H1): Standardization activities hinder research publication

Hypothesis 2 (H2): Standardization activities do not necessarily improve patenting.

3. METHOD AND RESULTS

3.1 Overview

We employed a method of data preparation and analysis similar to that used in earlier research (Newman, 2004; Shibata *et al*., 2010; Tashiro *et al*., 2013) and propose a procedure involving the vertical integration of exploration and exploitation (Table 1 and Appendix Fig. A.1).

[Insert Table 1 about here]

3.1.1 Data preparation

We used Web of Science and Thomson Innovation as data sources. Web of Science is an online database of academic papers that enables comprehensive citation searches. Thomson Innovation is a global patent database. We used MPEG as the analysis subject. This is a typical standardized technology, which means that we can clearly observe the influence of standardization on the knowledge space. We subtracted published works and patent filings from 1980 to 2014. We connect the keywords “mpeg” and “standardization” with a Boolean operator “AND” and prepare our two search strings as 1) (mpeg), and 2) (mpeg AND standardization). Following a key word search, we selected 6,560 papers and 42,904 patents from the databases for the search string (mpeg), and 1,535 papers and 7,347 patents for the search string (mpeg AND standardization).

3.1.2 Research Procedure

In the exploration phase, we used an analysis method similar to that used in earlier research (Newman, 2004; Shibata *et al*., 2010; Tashiro *et al*., 2013) for data preparation and BCCN. In the exploitation phase, we employed an additional procedure for this study (i.e., a new data processing treatment shown in Table 1).

3.2 Bibliographic Clustering of Citation Networks

After the clustering computation (Newman, 2004), 23 paper clusters and 111 patent clusters were obtained for (mpeg) and 14 paper clusters and 39 patent clusters for (mpeg AND standardization) (there was noise among the clusters, because "mpeg" is also used as a scientific phrase in chemistry, but we ignored these noisy clusters in the analysis). We only used clusters that had nodes above 100. We show the complete pictures of the clusters in Appendix Fig. B1, B2, B3, and B4. The size and major contents of the three largest relevant clusters are given in each figure.

3.2.1 Similarity between patent clusters and academic paper clusters

Shibata *et al*. (2010) and Iwami, Kogo, Tacoa, Mori, Kajikawa, and Sakata (2015) compared the bibliographic characteristics of clusters of patents and academic papers to observe technological similarities and potentially promising technology areas. To observe the similarities, we first selected important representative key words (Appendix Table C.1 and Appendix Table C.2) of each cluster by using the term frequency-inverse document frequency (TF-IDF) mutation method (Appendix D) in Layer 1 and Layer 2. In Figures 2 and 3, we show the heat maps for Layer 1 and Layer 2, having calculated the cosine similarities.

[Insert Figure 2 about here]

[Insert Figure 3 about here]

3.2.2 Similarities between Layer 1 and Layer 2

There is a method for exploiting the data shown in Figure 4 (Shibata *et al*., 2010), which is also useful for exploiting discovered knowledge. If the patent is extant and the publication of academic papers is insufficient (Area C), the industrial technology (patents) is leading the technology frontier more than the basic science (academic papers). Hence, we can see potential for the progressive development of basic research in this area of technology. This means that for researchers and institutions seeking research themes, this recently obtained information will be highly beneficial for exploitation. Conversely, if scholars have already published academic papers but patents are scarce (Area B), the R&D applied in this region has not been developed sufficiently; hence, there is a significant opportunity to obtain patents. Thus, the application of this method can affect the organization's search behavior in the knowledge space (Grant, 1996; Helfat, 1994; Mansfield, 1988; Nelson and Winter, 1982; Rosenkopf and Nerkar, 2001; Wade, 1996) and improve its ability to exploit the newly discovered knowledge. Moreover, science linkage is not evident in Area B and C, while it is evident in Area A. If the linkage pattern changes from the pattern A to B or from the pattern A to C, the science linkage is broken.

[Insert Figure 4 about here]

The comparison between Layer 1 and Layer 2 is the additional procedure that we undertake beyond the method used in prior studies (Shibata *et al*., 2010). In Layer 1 in Figure 2, there is consideration of 1) patents and 2) academic papers, and the key word and our search string is (mpeg). Similarly, in Layer 2, there is consideration of 1) patents, 2) academic papers, and 3) standardization, and our search string is (mpeg AND standardization). Comparing Layers 1 and 2, we can clearly see the influence of standardization on industrial technological development (patents) and basic research (academic papers). We use the difference in the heat maps to test the derived hypotheses.

4. DISCUSSION

4.1 Overview

Figure 2 and Figure 3 show the cosine similarities between clusters. The cosine similarity between patent cluster #x and academic paper cluster #y in Layer #z is denoted as:

$$\textit{Cosine similarity of Layer \#z between patent cluster \#x and academic paper cluster \#y} = \textit{Similarity}\ (x, y, z). \qquad (1)$$

For example, the cosine similarity in Layer 1 between patent cluster #1 and academic paper cluster #1

= $Similarity\ (1, 1, 1) = 0.436$, as shown in Figure 3.

In addition, the detected representative key words for each cluster are listed in Appendix Table C.1 and Table C.2. We denote the representative key words of each cluster as $Kw(x, y, z)$. When z = 1, it corresponds to Layer 1 and when z = 2, it corresponds to Layer 2. When y = 0, $Kw\ (x, 0, z)$ means the set of representative key words of the patent clusters in #x of Layer #z. When x = 0, $Kw(0, y, z)$ means the set of representative key words of the academic paper clusters in #y of Layer #z.

4.2 Characteristics of Layer 1

Figure 2 shows the relationship between industry technology (patents) and academic research (papers) under the search string (mpeg). For instance, technologies in the first patent cluster (#1) have terms in common with the academic papers in clusters #1 to #5. We express this as:

$$Kw(1,0,1) \cap Kw(0,i,1) \neq \phi, \qquad (2)$$

where i = 1 to 5.

This is seen in the fact that $Similarity\ (1, j, 1)\ (j = 1,2,3,4,5)$ is larger than 0.2. This reveals that technological information presented in papers is generally patented, and thus the result of fundamental research has been industrialized. Further, among combinations, $Kw(5,0,1)$ (Cluster 5) in the patents and $Kw(0,5,1)$ (Cluster 5) in in the academic papers have common representative key words of "watermark"3) (Appendix Table C.1). Then:

$$Kw(5,0,1) \cap Kw(0,5,1) \neq \phi. \qquad (3)$$

This is supported by the fact that $Similarity\ (5, 5, 1) = 0.378 > 0.2$. This means that academic research (papers) and industrial applications (patents) advance simultaneously regarding the technology of watermark, and there is a definite science linkage between them.

4.3 Characteristics of Layer 2

Figure 3 shows the relationship between industrial technology (patents) and academic research (papers) using the search string (mpeg AND standardization). It shows the cosine similarity between these two factors. The key words for each cluster are shown (Appendix Table C.2). Patent Cluster 7 has the representative key word "watermark," but there is no corresponding academic papers cluster with the term "watermark," as opposed to Layer 1. We denote this relationship as follows:

$$Kw(7,0,2) \cap Kw(0,i,2) = \phi, \qquad (4)$$

where i = 1 to 4.

This means that compared with industrial research (patents), fundamental research results (papers) are scarce in technologies relating to "watermark.". Therefore, the science linkage is low. While not hindering it, the standardization process may not improve basic research in the specified technology relating to "watermark.". In other words, standardization may not necessarily improve academic achievement in the form of academic papers. Conversely, the patenting process is not hindered by standardization.

4.4 Hypothesis validation

Hypothesis H1 is supported by the difference between Layer 1 and Layer 2. As shown in Figure 5, the standardization process may hinder academic achievement. In Layer 1, watermark technology represents clusters in both patents and academic papers, while in Layer 2, it represents clusters in patents, but does not in academic papers. This implies that the science linkage has altered between the two layers. This is confirmed by the fact that $Similarity\ (5, 5, 1) = 0.378$ (Layer 1) decreases to $Similarity\ (7, 0, 2) = 0\ to\ 0.1$ (Layer 2).

We can interpret this result as follows. In certain technology areas, standardization can suppress academic achievement. This is in accordance with the results observed in the research center in

---

[3] In this case, "watermark" means a digital watermark technology that inserts invisible signals into imaging data. This is used for authenticity validation, copyright tracking of digitized imaging, and detections of copyright infringement (Choi, Do, Choi, & Kim, 2010). It differs from conventional watermarks, which are translucent marks on prints or pictures.

Germany (Zi and Blind, 2015), and is confirmed in this study using the bibliographic method.

Hypothesis H2, on the other hand, is not supported. As shown by the difference between Layer 1 and Layer 2 in Figure 5, the standardization process may not necessarily hinder patenting. We think this is because the patented technology does not directly relate to the standardization, but develops around the standardized technologies of MPEG. Therefore, in this case there is a complementary relationship between standardization and patenting.

[Insert Figure 5 about here]

The knowledge structure observed in this study is shown in Figure 6. Knowledge of standardization is embedded in patent knowledge. Conversely, in academic paper knowledge, the effect of standardization creates a knowledge void. Total academic knowledge is divided into two areas, 1) (mpeg) and 2) (mpeg AND standardization). This means that it is difficult to simultaneously undertake standardization activities and academic research activities.

[Insert Figure 6 about here]

4.5 Theoretical Implications

The concept of exploration and exploitation outlined 20 years ago (March, 1991) was not based on the current ICT environment and ICT influence on knowledge acquisition. In the two decades since March's concept was presented, the theoretical assumptions and the notion of exploration and exploitation have implicitly changed as a result of the rapid development of ICT and the rapid decline in computation costs, as well as improvements in bibliographic data infrastructure (George *et al.*, 2014; George *et al.*, 2016). In this study, we wish to explicitly explain the implicit transition. The main change is that the optimal point of utility for knowledge acquisition between exploration and exploitation shifts substantially. We show in a vertical integration the two are now in a complementary relationship. Today, it has become possible using ICT.

We can find new knowledge (exploration) and apply the result to management strategy (exploitation) more efficiently and seamlessly than ever before because we can link the two actions in a complementary manner (i.e., in a vertically integrated manner). Bibliographic clustering reveals the knowledge structure shown in Figure 5 and identifies the knowledge region where standardization affects technology. This is an example of the proposed model representing the vertical integration of exploration and exploitation. From this perspective, we empirically find that every organization can more easily practice this strategy under certain conditions (Appendix Fig. A.1) and obtain knowledge and achieve innovation more efficiently. What is the difference between the previous strategy and proposed strategy? Previously, the relationship between exploration and exploitation was not necessarily complementary in many cases, therefore the strategy was difficult to implement. However, it is not necessarily empirically true in the current ICT environment that vertically integrated exploration and exploitation strategies are theoretically more difficult than horizontally integrated exploration and exploitation strategies.

4.6 Managerial implications

The findings of this study indicate that standardization activities require a lot of effort, and in some cases this consumes resources that scholars would otherwise use to produce academic papers. As a result, researchers cannot pursue academic paper preparation while involved in standardization activities. This result implies the need for behavioral change, for example, because of job security concerns. The same finding applies in the case of patenting (Huang, Feeney, and Welch, 2011). In the case of an academic researcher, a more senior faculty member has a greater propensity to patent than a younger faculty member because the senior faculty member may not be under the same pressure in terms of job security and thus may be more able to engage in patenting activities (Huang, Feeney, and Welch, 2011). We can apply the same theory to their respective attitudes toward standardization.

From the perspective of organizational economics, we can explain this result in terms of the multidimensional tasks problem (Holmstrom and Milgrom, 1991). For example, when production workers are responsible for producing a high volume of good-quality output, the volume is easy to measure but the quality is difficult to measure; thus the workers' attention is focused on volume (Holmstrom and Milgrom, 1991). In this context, we can easily measure patents under standardization, but we cannot measure academic papers since patents and academic papers are not necessarily compatible under the influence of standardization. As a result, the multidimensional task problem occurs in the MPEG case. To solve this problem, it is important that institutions design compensation contracts that reconcile the conflict of interest between principals (i.e., institutional interest) and agents (i.e., researchers' interest) (Baker, 1992; Kachelmeier and Williamson, 2010). This implies that there is a need for active policy support for the production of academic papers when researchers are involved in R&D where the influence of standardization is significant. Such policy support can include:

1) Compensation for academic papers produced under standardization
2) In the technology area where the standardization activities are essential, inclusion of standardization-related activities in the criteria used to determine eligibility for tenure.

Nevertheless, scholars have not addressed these issues in prior studies on the management of R&D organizations (Cabinet Office, 2016).

5. LIMITATIONS

Our study is not without limitations. We have examined the influence of standardization on the field of MPEG technology, but we should avoid any overgeneralization of these results. To try to find the same phenomenon elsewhere, it is necessary to identify the mechanism behind the observed results. Further, this study analyzed pooled data from specific periods. Hence, we cannot observe the dynamics of the relationship between patent cluster formation and academic paper cluster formation. If we could observe the dynamic development of each cluster, we could obtain more valuable information regarding standardization.

6. CONCLUSION

We examined the influence of image-digitizing technology (MPEG) developed in ISO and IEC by East Asian researchers. In MPEG technology, standardization plays an important role. This field is widely studied in relation to the application of artificial intelligence-related technology to the task of image recognition. Hence, our findings will have a significant impact on future work. Moreover, we examined the use of the bibliographic clustering methodology in exploring and exploiting the knowledge space. The results showed the possibility of overcoming the presumed limitations of exploration and exploitation beyond *bounded rationality* (Simon, 1979).

6.1 Theoretical Contribution

This study presents a new knowledge structure model. The structure model uses three-dimensional coordinates (Figure 1). By comparing the standardization-related and non-standardization-related layers, we addressed the influence of standardization on science linkage. In the case of digital watermark technology, standardization affects the science linkage, and a knowledge flow from patents to academic papers decreases. We observe that this field of study lacks academic papers, while patenting exists under standardization.

6.2 Managerial Contribution

We show that a combination of exploration and exploitation is possible in practice under the current ICT and big data environment. As a result of *exploration*, we found the knowledge space and the technology region where *exploitation* is most promising. This means that a vertical integration between *exploration* and *exploitation* is possible, rather than merely the conventional horizontal integration between the two. In the conventional horizontal integration concept, it is considered important to find a balance between exploration and exploitation, while a vertical integration, which we propose, allows for maximization of both at the same time.

Moreover, we applied bibliographic clustering analysis to explore organizational economics issues in R&D organizations and highlight the fact that research publication and standardization activities are not autonomously optimized (Zi and Blind, 2015). We explain this phenomenon from the viewpoint of the multidimensional task problem (Holmstrom and Milgrom, 1991). This is especially true in the case of individual researchers in institutions where academic achievement is important. To accommodate these two elements, it is necessary to implement a management system that reconciles the interests of both principals and agents, and provides incentives for academic achievement under conditions of standardization.

APPENDICES

APPENDIX A

[Insert Fig. A.1 about here]

APPENDIX B

[Insert Fig. B.1, B.2, B.3, and B.4 about here]

APPENDIX C

[Insert Table C.1 about here]

[Insert Table C.2 about here]

APPENDIX D: Calculation of Cosine Similarity between Clusters

Using the key words in each cluster, we compared the bibliographic similarities between i) patent clusters and ii) academic paper clusters to reveal the technological analogy. To measure the similarity, we used the cosine similarity formula. This is given by the following equation:

$$Cosine\ similarity = \frac{\overrightarrow{v1} \cdot \overrightarrow{v2}}{|\overrightarrow{v1}| \cdot |\overrightarrow{v2}|},$$

where $\overrightarrow{v1}$ and $\overrightarrow{v2}$ are vectors of word frequency in each cluster, for example $\overrightarrow{v1}$ = (frequency of word 1, frequency of word 2,…, frequency of word n) = (f_1, f_2,…, f_n).

This cosine similarity ranges in value from 0 to 1. When the similarity is 1, these two vectors of cluster are identical and the two clusters are similar. Conversely, when the similarity is 0, these two vectors of clusters are completely different sets of words, and therefore the two clusters are not similar. We depict the calculated values in the heat map, where a deep color represents a high degree of similarity and white areas indicate the absence of a relationship. The heat map shows 1) the most related clusters and 2) the most unrelated clusters.

REFERENCES

Acs, Z. J., and Audresch, D. B. 1989. Patents as a measure of innovative activity. ***Kyklos***, 42(2): 171–180.

Adai, A. T., Date, S. V., Wieland, S., and Marcotte, E. M. 2004. LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks. ***Journal of Molecular Biology***, 340(1): 179–190.

Albusac, J., Castro-Schez, J. J., Lopez-Lopez, L. M., Vallejo, D., and Jimenez-Linares, L. 2009. A supervised learning approach to automate the acquisition of knowledge in surveillance systems. ***Journal Signal Processing***, 89(12): 2400–2414.

Alexander, C. 1964. ***Notes toward a synthesis of form***. Cambridge, MA: Harvard University Press.

Amato, G., Savino, P., and Magionami, B. 2008. ***Use of weighted visual terms and machine learning techniques for image content recognition relying on mpeg-7 visual descriptors.*** MS '08 Proceedings of the 2nd ACM workshop on multimedia semantics: 60–63.

Arthur, W. B. 1989. Competing technologies, increasing returns, and lock-in by historical events. ***Economic Journal***, 99: 116–131.

Baker, G. 1992. Incentive contracts and performance measurement. ***Journal of Political Economy***, 100(3): 598–614.

Blind, K., and Gauch, S. 2009. Research and standardisation in nanotechnology: Evidence from Germany. ***The Journal of Technology Transfer***, 34(3): 320–342.

Borner, K., Chen, C., and Boyack, K. W. 2003. Visualizing knowledge domains. ***Annual Review of Information Science and Technology***, 37(1): 179–255.

Cabinet Office. 2016. ***The Fifth Science and Technology Basic Plan***. Tokyo: Japanese government.

Carpenter, M., and Narin, F. 1983.Validation study: Patent citations as indicators of science and foreign dependence. ***World Patent Information***, 5(3): 180–185.

Carpenter, M., Narin, F., and Woolf, P. 1981. Citation rates to technologically important patents. ***World Patent Information***, 3(4): 160–163.

Choi,D., Do,H., Choi,H., and Kim,T. 2010. A blind MPEG-2 video watermarking robust to camcorder recording. ***Signal Processing***,90(4):1327-1332.

Cohen, W.M., Nelson, R.R., and Walsh, J.P. 2002. Links and impacts: The influence of public research on industrial R&D. ***Management Science***, 48(1): 1–23.

Colbert, A., Yee, N., and George, G. 2016.The digital workforce and the workplace of the future. ***Academy of Management Journal***, 59(3): 731–739.

Dahlander, L., O'Mahony, S., and Gann, D.M. 2016. One foot in, one foot out: How does individuals' external search breadth affect innovation outcomes? ***Strategic Management Journal***, 37(2): 280–302.

David, P. A. 1985. Clio and the economics of QWERTY. ***American Economic Review***, 75(2): 332–337.

Demirkan, I., Deeds, D. L., and Demirkan, S. 2013. Exploring the role of network characteristics, knowledge quality, and inertia on the evolution of scientific networks. ***Journal of Management***, 39(6): 1462–1489.

Demirkan, I., and Demirkan, S. 2012. Network characteristics and patenting in biotechnology: 1990–2006. ***Journal of Management***, 38(6), 1892–1927.

Dewett, T., and Denisi, A. S. 2004. Exploring scholarly reputation: it's more than just productivity. ***Scientometrics***, 60(2): 249–272.

Edler, J., Georghiou, L., Blind, K., and Uyarra, E. 2012. Evaluating the demand side: New challenges for evaluation. ***Research Evaluation***, 21(1): 33–47.

Egyedi, T. M., and Sherif, M. H. 2010. Standards dynamics through an innovation lens: Next-generation ethernet networks. ***IEEE Communications Magazine***, 48(10): 166–171.

Enkel, E., Heil, S., Hengstler, M., and Wirth, H. 2016. Exploratory and exploitative innovation: To what extent do the dimensions of individual level absorptive capacity contribute? ***Technovation*** (in press). http://dx.doi.org/10.1016/j.technovation.2016.08.002

Felten, E. 2016. ***Preparing for the future of artificial intelligence.*** The White House, Washington DC. Accessed at https//www.whitehouse.gov/blog/2016/05/03/ preparing-future-artificial-intelligence (July 1, 2016).

Fernández-Escribano, G., Kalva, H., Cuenca, P., and Orozco-Barbosa, L. 2006. ***A very low complexity MPEG-2 to H.264 transcoding using machine learning.*** Proceeding of the ACM Multimedia 2006:931–940.

Fernández, G., Kalva, H., Cuenca, P., and Orozco-Barbosa, L. 2007. A first approach to speeding-up the inter mode selection in MPEG-2/H.264 transcoders using machine learning. ***Multimedia Tools and Applications***, 35(2): 225–240.

Frantzi, K., Ananiadou, S., and Mima, H. 2000. Automatic recognition of multi-word terms: The C-value /NC-value method. ***International Journal on Digital Libraries***, 3(2): 115–130.

Funk, J. L., and Luo, J. 2015. Open standards, vertical disintegration and entrepreneurial opportunities: How vertically-specialized firms entered the U.S. semiconductor industry. ***Technovation***, 45–46: 52–62.

Garfield, E. 1955. Citation indexes for science: A new dimension in documentation through association of ideas. ***Science***, 122: 108–111.

George, G., Haas, M. R., and Pentland, A. 2014. Big data and management. ***Academy of Management***

Journal, 57(2): 321–326.

George, G., Osinga, E.C., Lavie, D., and Scott, B.A. 2016. Big data and data science methods for management research. ***Academy of Management Journal***,59(5):1493-1507.

Grant, R.M. 1996. Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. ***Organization Science***, 7(4): 375–387.

Griliches, Z. 1984. ***R&D, patents, and productivity***. Chicago, IL: University of Chicago Press.

Griliches, Z. 1990. Patent statistics as economic indicators: A survey. ***Journal of Economic Literature***, 28(4):1661–1707.

He, Z. L., and Wong, P. K. 2004. Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. ***Organization Science,*** 15(4): 481–494.

Helfat, C.E. 1994. Evolutionary trajectories in petroleum firm R&D. ***Management Science***, 40(12): 1720–1747.

Holmstrom, B., and Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. ***Journal of Law, Economics & Organization***, 7: 24–52.

Huang, K.G., and Murray, F.E. 2009. Does patent strategy shape the long-run supply of public knowledge? Evidence from human genetics. ***Academy of Management Journal***, 52(6): 1193–1221.

Huang, W., Feeney, M. K., and Welch, E. W. 2011. Organizational and individual determinants of patent production of academic scientists and engineers in the United States. ***Science and Public Policy***, 38(6): 463–479.

Iwami, S., Kogo, K., Tacoa, F., Mori, J., Kajikawa, Y., and Sakata, I. 2015. ***Cross-disciplinary methodology for detection of collaborative and competitive candidates***. Proceedings of the 24th international conference for management of technology (IAMOT 2015), South Africa.

Jaffe, A.B. 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. ***American Economic Review***, 76(5): 984–1001.

Jakobs, K., Procter, R., and Williams, R. 2001. Standardisation, innovation and implementation of information technology. ***Computers and Networks in the Age of Globalization***, 57: 201–217.

Kachelmeier, S. J., Reichert, B. E., and Williamson, M. G. 2008. Measuring and motivating quantity, creativity, or both. ***Journal of Accounting Research***, 46(2): 341–373.

Kachelmeier, S. J., and Williamson, M. G. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. ***The Accounting Review***, 85(5): 1669–1691.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. ***Econometrica***, 47(2): 263–292.

Katila, R. 2002. New product search over time: Past ideas in their prime? ***Academy of Management Journal***, 45(5): 995–1010.

Katila, R., and Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. ***Academy of Management Journal***, 45(6): 1183–1194.

Laursen, K., and Salter, A. 2006. Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. ***Strategic Management Journal***, 27(2): 131–150.

Levitt, B., and March, J. G. 1988. Organizational learning. ***Annual Review of Sociology***, 14: 319–340.

Leydesdorff, L. 1995. ***The challenge of scientometrics: The development, measurement, and self-organization of scientific communications***. Amsterdam, Netherlands: DSWO Press.

Mansfield, E. 1988. The speed and cost of industrial innovation in Japan and the United States: External vs. internal technology. ***Management Science***, 34(10): 1157–1168.

March, J. G. 1991. Exploration and exploitation in organizational learning. ***Organization Science***, 2(1): 71–87.

McMillan, G. S., Narin, F., and Deeds, D. L. 2000. An analysis of the critical role of public science in innovation: The case of biotechnology. ***Research Policy***, 29(1): 1-8.

Milgrom, P.R., and Roberts, J. 1992. ***Economics, organization and management***. New Jersey:

Prentice-Hall.

Nelson, R.R., and Winter, S.G. 1982. ***An evolutionary theory of economic change.*** Cambridge, MA: Harvard University Press.

Newman, M.E.J. 2004. Fast algorithm for detecting community structure in networks. ***Physical Review E***, 69(6).

Oliver, A. L. 2004. Biotechnology entrepreneurial scientists and their collaborations. ***Research Policy***, 33(4): 583–597.

Oliver, A. L., and Liebeskind, J. P. 1997. Three levels of networking for sourcing intellectual capital in biotechnology. ***International Studies in Management & Organization***, 27(4): 76–103.

Oyer, P. 2006. Initial labor market conditions and long-term outcomes for economists. ***Journal of Economic Perspectives***, 20(3): 143–160.

Rosenkopf, L., and Nerkar, A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. ***Strategic Management Journal***, 22(4): 287–306.

Salter, A. J., and Martin B. R. 2001. The economic benefits of publicly funded basic research: A critical review. ***Research Policy***, 30(3): 509–532.

Sherif, M. H. 2001. A framework for standardization in telecommunications and information technology. ***IEEE Communications Magazine***, 39(4): 94–100.

Shibata, N., Kajikawa, Y., and Sakata, I. 2010. Extracting the commercialization gap between science and technology – Case study of a solar cell. ***Technological Forecasting and Social Change***, 77(7): 1147–1155.

Shibata, N., Kajikawa, Y., Takeda, Y., and Matsushima, K. 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. ***Technovation***, 28(11): 758–775.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. 2016. Mastering the game of Go with deep neural networks and tree search. ***Nature***, 529: 484–489.

Simon, H. A. 1979. Rational decision making in business organizations. ***American Economic Review***, 69(4): 493–513.

Tamura, S. 2012. Effects of integrating patents and standards on intellectual property management and corporate innovativeness in Japanese electric machine corporations. ***International Journal of Technology Management***, 59(3/4): 180–202.

Tamura, S. 2013. Generic definition of standardization and the correlation between innovation and standardization in corporate intellectual property activities. ***Science and Public Policy***, 40(2): 143–156.

Tamura, S. 2016. A new intellectual property metric for standardization activities. ***Technovation***, 48–49: 87–98.

Tashiro, A., Tashiro, H., Iwami, S., and Sakata, I. 2013. ***Bibliometric analysis of net zero energy building.*** Proceedings of the PICMET ’13 Conference, Vancouver, Canada.

Tassey, G. 2003. ***Method for assessing the economic impacts of government R&D***. Gaithersburg, MD: National Institute of Standards & Technology.

Tortoriello, M. 2015. The social underpinnings of absorptive capacity: The moderating effects of structural holes on innovation generation based on external knowledge. ***Strategic Management Journal***, 36(4): 586–597.

Trajtenberg, M. 1990a. A penny for your quotes: patent citations and the value of innovations. ***The RAND Journal of Economics***, 21(1): 172–187.

Trajtenberg, M. 1990b. ***Economic analysis of product innovation: The case of CT scanners.*** Cambridge, MA: Harvard University Press.

Trajtenberg, M., Henderson, R., and Jaffe, A.B. 1997. University versus corporate patents: A window on the basicness of invention. ***Economics of Innovation and New Technology***, 5(1): 19–50.

Looy, B.V., Zimmermann, E., Veugelers, R., Verbeek, A., Mello, J., and Debackere, K. 2003. Do science-technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domains. ***Scientometrics***, 57(3): 355–367.

von Hippel, E. 1994. “Sticky information” and the locus of problem solving: implications for innovation. ***Management Science***, 40(4): 429–439.

Wade, J. 1996. A community-level analysis of sources and rates of technological variation in the microprocessor market. ***Academy of Management Journal***, 39(5): 1218–1244.

Yasuda, H. 1989. Standardization activities on multimedia coding in ISO. ***Signal Processing: Image Communication***, 1(1): 3–16.

Zi, K., and Blind, K. 2015. Researchers’ participation in standardisation: A case study from a public research institute in Germany. ***The Journal of Technology Transfer***, 40(2): 346–360.

Zucker, L. G., Darby, M.R., and Armstrong, J. S. 2002. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. ***Management Science***, 48(1): 138–153.

Tables and Figures:

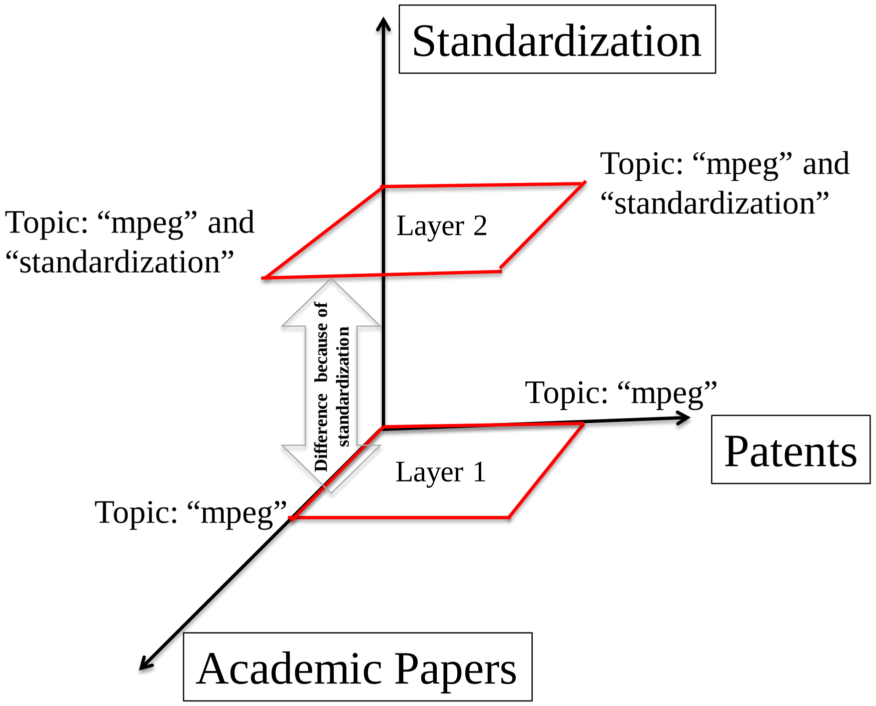
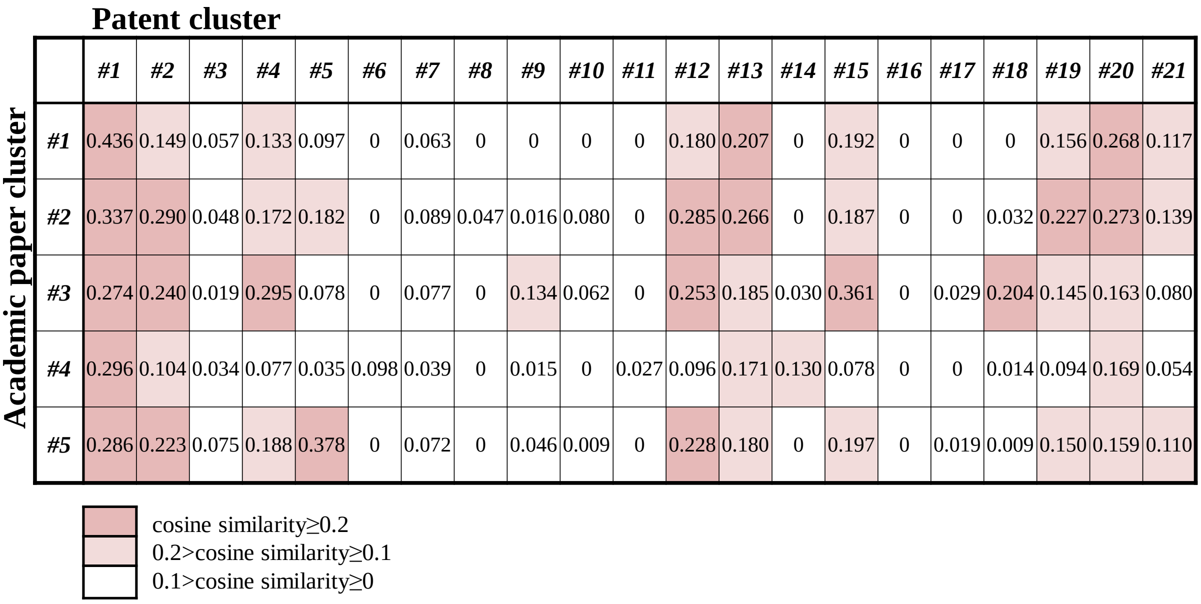


Figure 1. Three-dimensional knowledge space structure model



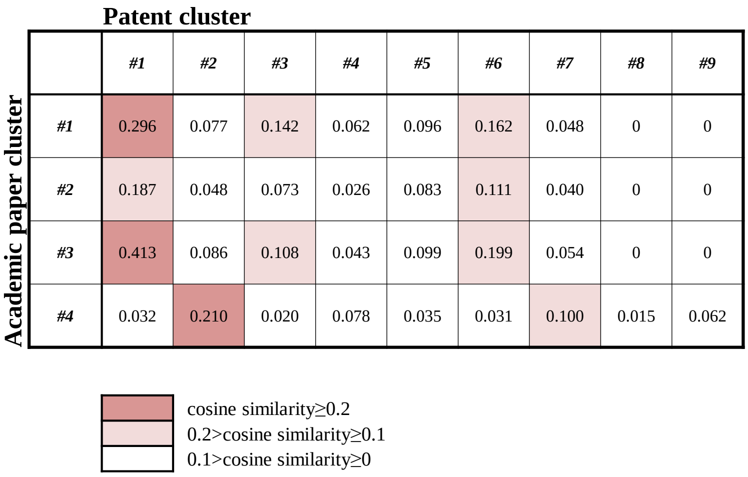
Patent cluster

Academic paper cluster

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21
#1	0.436	0.149	0.057	0.133	0.097	0	0.063	0	0	0	0	0.180	0.207	0	0.192	0	0	0	0.156	0.268	0.117
#2	0.337	0.290	0.048	0.172	0.182	0	0.089	0.047	0.016	0.080	0	0.285	0.266	0	0.187	0	0	0.032	0.227	0.273	0.139
#3	0.274	0.240	0.019	0.295	0.078	0	0.077	0	0.134	0.062	0	0.253	0.185	0.030	0.361	0	0.029	0.204	0.145	0.163	0.080
#4	0.296	0.104	0.034	0.077	0.035	0.098	0.039	0	0.015	0	0.027	0.096	0.171	0.130	0.078	0	0	0.014	0.094	0.169	0.054
#5	0.286	0.223	0.075	0.188	0.378	0	0.072	0	0.046	0.009	0	0.228	0.180	0	0.197	0	0.019	0.009	0.150	0.159	0.110

cosine similarity≥0.2
0.2>cosine similarity≥0.1
0.1>cosine similarity≥0

Figure 2. Heat map (Layer 1): cosine similarities between patent clusters and academic paper clusters



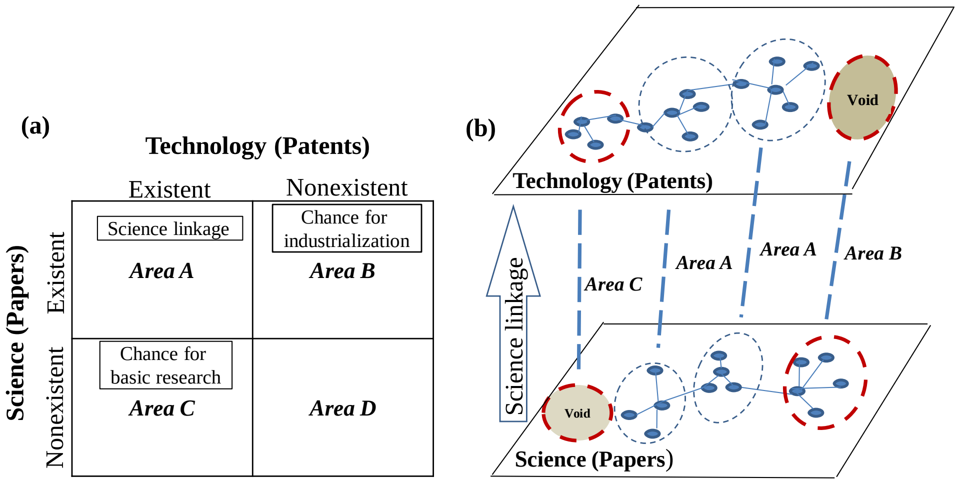
Patent cluster

Academic paper cluster

	#1	#2	#3	#4	#5	#6	#7	#8	#9
#1	0.296	0.077	0.142	0.062	0.096	0.162	0.048	0	0
#2	0.187	0.048	0.073	0.026	0.083	0.111	0.040	0	0
#3	0.413	0.086	0.108	0.043	0.099	0.199	0.054	0	0
#4	0.032	0.210	0.020	0.078	0.035	0.031	0.100	0.015	0.062

cosine similarity≥0.2
0.2>cosine similarity≥0.1
0.1>cosine similarity≥0

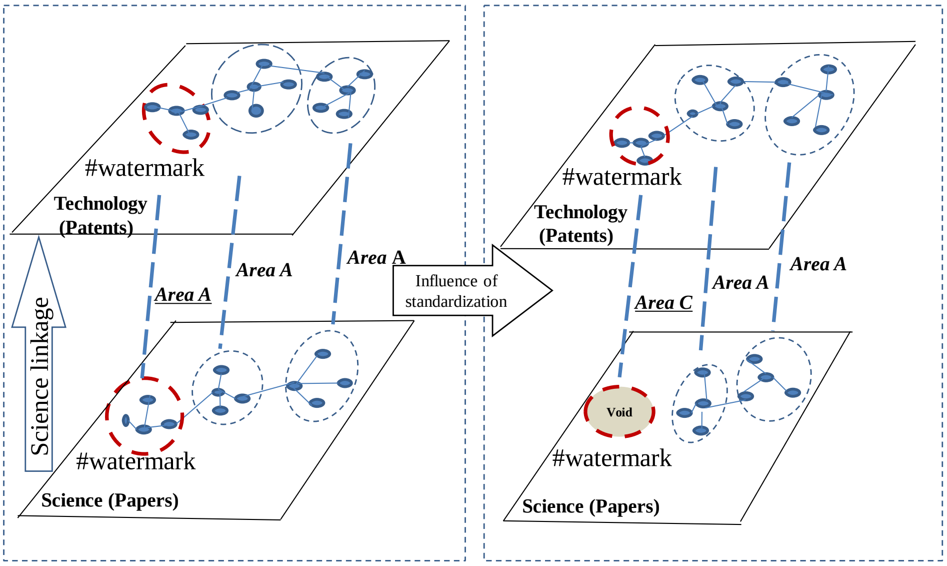
Figure 3. Heat map (Layer 2): cosine similarities between patent clusters and academic paper clusters (with standardization)



		Technology (Patents)	
		Existent	Nonexistent
Science (Papers)	Existent	Science linkage ***Area A***	Chance for industrialization ***Area B***
	Nonexistent	Chance for basic research ***Area C***	***Area D***

Reference: Shibata *et al.* (2010)
Figures are made by the authors.

Figure 4. Relationship between science and technology:
(a) Four types of relationship; (b) Gap between science and technology



Layer 1 **Layer 2**

Figure 5. Change in science linkage arising from standardization

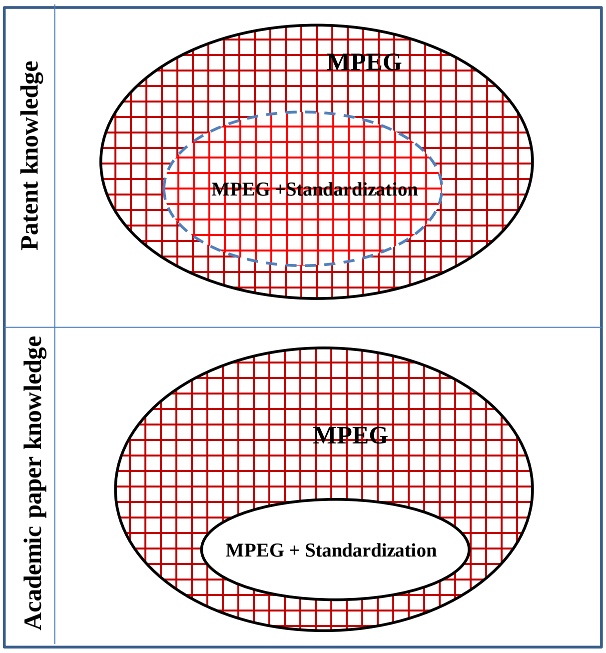


Figure 6. Corresponding relationship between patents and academic papers in Layer 2

Table 1. Data preparation and processing procedure

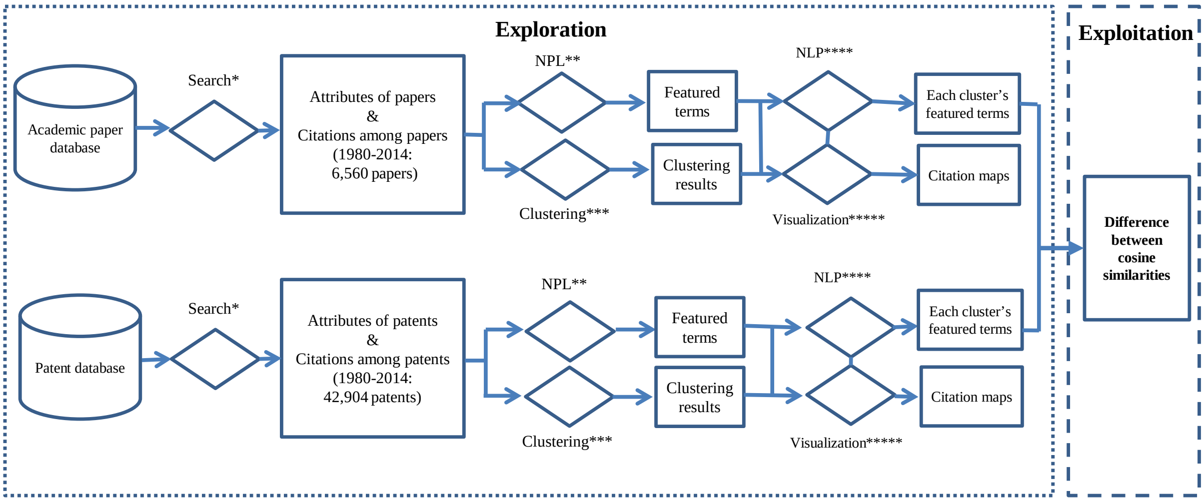
	Data preparation	Data processing	
		Bibliographic Clustering of Citation Networks (BCCN) (Newman, 2004; Shibata et al., 2010; Tashiro et al., 2013)	**New additional treatment of this study**
Exploration	We select the relevant academic papers and patents for clustering using the key words: 1) "mpeg" 2) "mpeg" and "standardization" We used the following bibliographic information sources from the Thomson Reuters database for the subtraction process: 1) Title information in Thomson Innovation (patents) 2) Title information in Web of Science (academic papers).	1) Subtracting the patents and academic papers from patent and journal databases using intended key words (this process selects targeted patents and journal papers) 2) Clustering patents and academic papers through unsupervised learning in terms of citation networks (this process clusters patents and papers with similar bibliographic characteristics) (The Academic Landscape* system is used for the computation of clustering and consequent comparison between clusters.)	
Exploitation			1) Chart the heat map of cosine similarities between i) patents and ii) academic papers (Layer 1). 2) Chart the heat map of cosine similarities between i) patents and ii) academic papers (Layer 2). 3) Find the specific technologies whose cosine similarity between patent clusters and academic paper clusters differs between Layer 1 and Layer 2. 4) Observe the science linkage of specified technologies and find a specific technology region.

Note: * Copyright © 2010-2013 Innovation Policy Research Center, The University of Tokyo, and Copyright © 2012-2013 Kajikawa laboratory, Graduate School of Innovation Management, Tokyo Institute of Technology

Appendices:

Appendix A

Fig. A.1 Detailed flow chart: Vertical integration of exploration and exploitation



Note: * In this case, the queries are 1) "mpeg" or 2) "mpeg" and "standardization."
** NC-Value methods (Frantzi, Ananiadou, and Mima, 2000).
*** The clustering algorithm (Newman, 2004).
**** TF-IDF method.
***** The Large Graph Layout (LGL) visualizing engine (Adai *et al.*, 2004).

Appendix B

Fig. B.1　Clusters of academic papers (mpeg)

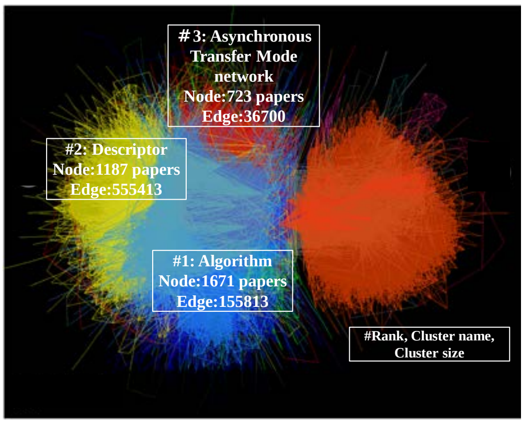


Fig. B.2　Clusters of academic papers
(mpeg AND standardization)

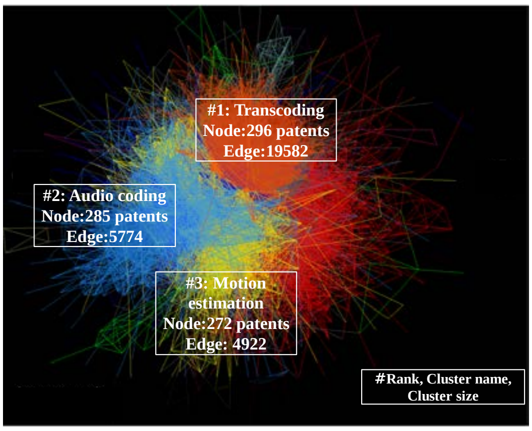


Fig. B.3 Clusters of patents (mpeg)

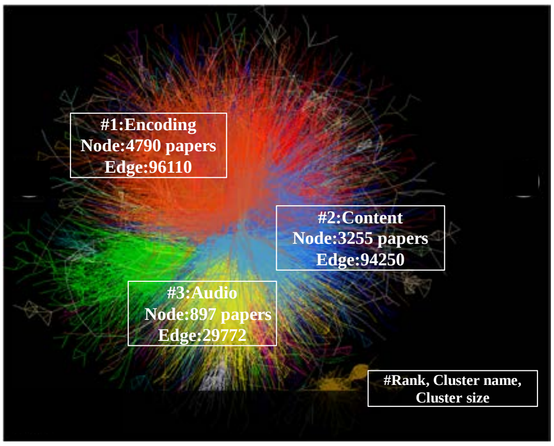
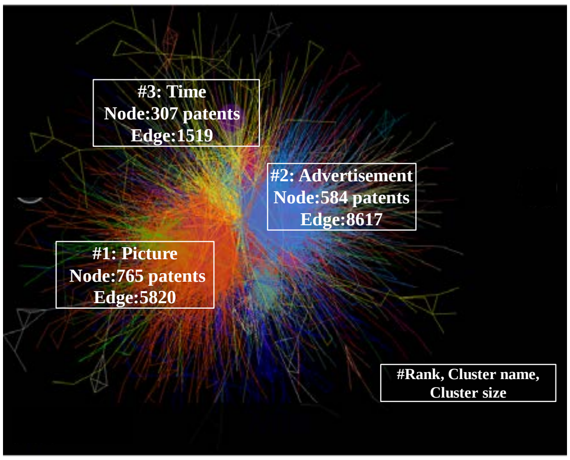


Fig. B.4 Clusters of patents (mpeg AND standardization)



Appendix C

Table C.1 Keywords in academic paper clusters and patent clusters (Layer 1)

Academic paper cluster	Top TF-IDF terms
#1	coding, video, bit, algorithm, video coding
#2	video, object, image, motion, descriptor
#3	video, traffic, network, atm, error
#4	motion estimation, estimation, search, motion, video
#5	watermarking, video, watermark, quality, watermarking scheme

Patent cluster	Top TF-IDF terms
#1	video, block, frame, encoding, image
#2	content, medium, video, program, user
#3	audio, medium, file, content, player
#4	packet, stream, video, transport, data
#5	content, watermark, digital, file, medium
#6	memory, memory device, memory cell, flash, flash memory
#7	information storage medium e.g, information storage medium, specific unit, information storage, dvd ram
#8	touch, user, touch screen, sensor, electronic
#9	packet, broadcast, digital, stream, data
#10	content, sponsor, communication facility, mobile communication facility, user
#11	card, electronic, case, electronic device, cover
#12	network, audio, image, video, remote
#13	stereoscopic, dimensional, video, image, picture
#14	power, charging, wireless power, power transmission, battery
#15	packet, video, stream, transmission, frame
#16	interferometric, light, interferometric modulator, microelectromechanical, modulator
#17	caption, caption service, transmitting digital broadcast, transmitting digital broadcast signal, closed caption service
#18	wireless, network, communication, node, access
#19	volume descriptor, recording, volume descriptor sequence, descriptor sequence, descriptor
#20	encoding, image, picture, frame, shot
#21	image, light, organic light emitting display, emitting display, light emitting display

Table C.2 Keywords in academic paper clusters and patent clusters (Layer 2)

Academic paper cluster	**Top TF-IDF terms**
#1	**mode, video, h.264/avc, transcoding, rate**
#2	**audio, mdct, audio coding, transform, dct**
#3	**object, segmentation, motion, motion estimation, block**
#4	**descriptor, retrieval, content, multimedia, shape**

Patent cluster	**Top TF-IDF terms**
#1	**video, block, picture, motion, frame**
#2	**content, program, network, video, advertisement**
#3	**video, packet, recording, stream, time**
#4	**encrypted, content, key, packet, encryption**
#5	**dimensional, video, image, picture, stereoscopic**
#6	**artifact, block, filtering, pixel, video**
#7	**watermark, content, document, blanking interval, blanking**
#8	**enhanced data, trellis, vestigial, traffic information data, traffic information**
#9	**connector, interface, medium, file, card**