LEGAL PROTECTION OF MOBILE P2P DATABASES

Olli Pitkänen, Perttu Virtanen, Mikko Välimäki
Helsinki Institute for Information Technology
P.O.Box 9800, 02015 HUT
Finland

ABSTRACT

The paper studies the legal protection of databases and especially the European *database sui generis right* from the mobile business perspective. First, we discuss about arguments in favor and against legal protection with respect to current databases. Next, we describe the future model of mobile peer-to-peer databases. Finally, we analyze how the database right suits to the future databases and the mobile business. We conclude that the definition of database in the EU directive has its shortcomings. In some cases mobile businesses should implement digital rights management into their database services if usage control is needed. We also find that the general idea of protecting substantial investments in databases will be valid in the mobile business.

KEY WORDS

Intellectual Property Law, Database Protection, Information Technology, Peer-to-Peer Technology (P2P), Digital Rights Management

1. INTRODUCTION

Many business opportunities on the mobile networks rely on information systems that are built on top of databases. In general, more and more information is stored as data in databases. Therefore, databases form a crucial tool in the development of the mobile business. On the other hand, peer-to-peer (P2P) networks enable interesting new solutions to build distributed databases.

In this paper, we concentrate on the *sui generis* right that is specifically meant to cover databases and their contents. We recognize, however, that the database systems can be partially protected also by many other kinds of intellectual property rights, like copyright and patents.

Let us first define a couple of basic concepts: data and information:

Data are numbers, characters, images, or other method of recording, in a form which can be assessed by a human or especially input into a computer, stored and processed there, or transmitted on some digital channel. Computers nearly always represent data in binary. Data on its own has no meaning. People or computers can find patterns in data to perceive information.

Information is stimuli that have meaning in some context for its receiver. Information can be converted into data and passed on to another receiver. Relative to the computer, we can say that: Information is made into data, put into the computer where it is stored and processed as data, and then put out as data in some form that can be perceived as information. [11]

For example number 1234.56 is data. In contrast, "your bank balance is $1234.56" is information. Another way to illustrate how data differs from physical medium and information is to start from the user point of view and imagine text in a piece of paper. If the text is looked close enough – using a microscope or a magnifying class, for example – one can see the details of the surface of the paper and ink on it. That is physical medium. When the viewer moves away from the paper, single characters or letters can be seen. That is data. When the distance increases further, one can see words, sentences, and paragraphs and starts to understand what the text says. That is information.

2. LEGAL AND TECHNICAL CONCEPTS OF DATABASE

The word database is ambiguous. Especially, a 'database' in information technology and a 'database' in legal context are not necessarily the same. In the following we discuss about differences in the legal and technical concepts of database.

2.1 Legal Databases

Copyright protects original, expressed, creative works. In general, an idea is not copyrightable, but on certain conditions it can be *patentable* or it may be possible for example to claim it as a *trade secret*. (See Figure 1.) The expression of an idea may be copyrighted. On the other hand, if the same idea is expressed in different, independent ways, each of those expressions can be a copyrighted work of its own and they do not infringe each other. The physical embodiments or the copies of copyrighted ex-

pressions can be for instance sold without assigning copyright. [5], [8], [9], [10]

Object	Examples	Means of protection
Information	Abstract ideas, facts, knowledge, wisdom	No legal rights
	Ideas reduced to practice	Possibly patents, trade secrets, etc.
	Expression of ideas, creativity, etc.	Possibly copyright, trademarks, trade secrets, etc.
Data	Representation e.g. in binary form	No legal rights, but possibly technical protection, e.g. encryption
Physical medium	Embodiment	Possibly property rights, technical protection

Figure 1. Levels of abstraction related to immaterial objects and their protection

A database can include copyrighted works and even a database as a whole can be copyrighted if it is original enough. However, most databases are not copyrightable and their content is not copyrighted either. Yet, the making of databases requires the investment of considerable human, technical and financial resources while such databases can be copied or accessed at minimal cost. Therefore some kind of protection for databases is needed.

European Union has adopted a directive concerning the legal protection of databases. It recognizes the possibility of copyrighting a database but also defines a neighboring right, a specific *sui generis database right*. [4] Several other countries are considering similar statutes. In the USA, a number of bills have been introduced in relation to database protection, but no statutes have been passed so far. [4], [6]

All the EU member countries need to have implemented the directive. However, they have had the liberty to implement it in their own ways. Therefore the database legislation differs slightly within the European Union. For example, in the United Kingdom, the legislator has chosen to include the definition of a database in the statute quite directly from the EU directive: "'database' means a collection of independent works, data or other materials which (a) are arranged in a systematic or methodical way, and (b) are individually accessible by electronic or other means" and a "property right ('database right') subsists […] in a database if there has been a substantial investment in obtaining, verifying or presenting the contents of the database." [3] In Finland, on the other hand, the legislator has chosen not to specifically define database in the statute, but to declare only that the sui generis right requires a substantial investment in obtaining, verifying or presenting the contents of the database. [10]

According to the database directive, the term 'database' means a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means. Databases should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data. This means that a recording or an audiovisual, cinematographic, literary or musical work as such is not a database. On the other hand, not all the databases that fulfill this definition gain database protection. It is namely further required that in order to get the sui generis right in a database, it must show that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents. [4]

Contents / The Whole	not original	original
no substantial investments	No legal protection	The whole not legally protected, but copyrighted contents
substantial investment	Database sui generis right	Sui generis right + copyrighted contents
original	Copyrighted as a whole	Copyrighted both as a whole and contents

Figure 2. Copyright and the Sui Generis Right in a Database as a Whole and in Contents.

Clearly, "works, data or other material" in the definition of database are quite troublesome. The directive is obviously trying to state that databases can include many kinds of information, copyrighted works as well as other sets of information. The wording, however, is quite unsuccessful. 'Data' here do not refer to methods of recording as defined above, but rather to information. 'Material' on the other hand probably refers to *im*material items. Therefore this part of the definition does not help us very much.

2.2 Technical Databases

From the technical point of view, a database system in a computer consists of several components. There is a collection of data and a collection of programs to access the data. According to Korth and Silberschatz, a major purpose of a database system is to provide users with an abstract view of the data. That is, the system hides certain details of how the data is stored and maintained. This is accomplished by defining three levels of abstraction at which the database may be viewed: the physical level, the conceptual or logical level, and the view level. Physical level describes how a record is stored. Logical level describes data stored in database, and the relationships among the data. On view level, application programs hide details of data types. Views can hide information for security purposes. There can be different views for each user based on for example users' needs, rights, and security requirements. [7] It seems that many database systems perform this task in such an excellent way that most users cannot make distinction between the three levels of abstraction. Instead they think that the view they see is the actual database. Unfortunately, the legislators do not seem be able to avoid that confusion. This makes the legal analysis quite difficult. What is the subject matter of the database protection? Is it the view a user sees or the actual data stored on the physical level or something in between?

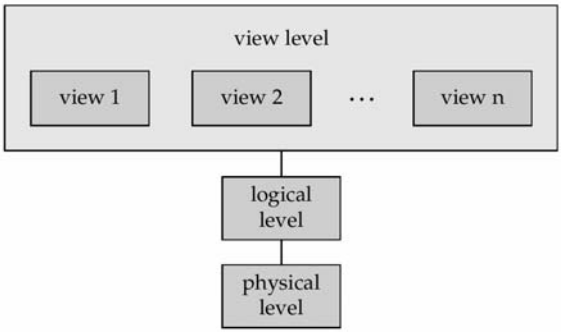

Figure 3. The three levels of a database system. [7]

Let us get back to the legal definition of 'database' in the directive: a collection of independent works, data or other materials *arranged in a systematic or methodical* way and individually accessible by electronic or other means. We are mostly concerned about the requirement of systematic or methodical arrangement. Let us consider an example. Suppose a group of biologists makes a detailed catalog of natural resources of wildlife and game in a particular area. It takes months to collect, store, and verify the data. The outcome is valuable as the inventory can be used in many studies. However, such a list of natural resources is not necessarily "arranged in a systematic or methodical way". To be valuable, the list does not need to be even in alphabetical order. It can be just the Latin names of species in a random order but a user can still analyze the information with a computer. Is that a systematic or methodical way? Hardly. If the sui generis right requires more than trivial arrangement of data then valuable lists – even if they have needed substantial investments – do not gain the right. On the other hand, the directive does not require any qualitative or quantitative criteria for arrangement.

In a computer-based system, databases are typically arranged by attaching an index to them. For example, data items can be stored into a database in whatever order they arrive, but a constantly updated index is used to keep the data items in order. This can be done fully automatically so that a user does not see indices and the indexing process at all. The actual data can be completely unarranged. Yet, a user can make queries and the database system software shows results arranged as the user wants. This representation does not necessarily have anything to do with the actual arrangement or unarrangement of data in the database. An index is meant to increase the performance of a database system so that frequent queries can be completed rapidly. However, indices are usually not mandatory. A database system can be fully functional, only somewhat slower, without indices. If there are no indices or no index is useful for a certain query, the system must at worst go through each data item to decide whether it matches the query. This takes computing power and time, but it does not affect the results.

3. PROTECTED SUBJECT MATTER – WHAT IS VALUABLE IN DATABASES?

The sui generis database right requires substantial investment in obtaining, verifying or presenting the contents of the database. Certainly individual data items can be valuable, but as discussed earlier, they should not be protected as such in general. Instead, it can make a lot of sense to protect large investments that are needed to obtain, to verify, and to present the contents of the database. But as proposed above, it is possible to interpret the EU directive in a way that the sui generis right does not protect investments as such but their actual results or value to users. From this point of view, the arrangement of a *Sui Generis* database is not essential; it needs not inevitably be included in the definition of 'database', although a significant investment in arranging data can help to achieve the sui generis right. Of course lowering the threshold for protection in this respect requires further careful consideration to check the dangers of encroachment.

In Figure 1, we have illustrated how the levels of abstraction related to intangible objects affect legal protection. The question arises, does database protection fit into the picture and on which level it would be. Obviously, the

database sui generis right has a significant extra dimension – investment – that is not shown in the Figure 3. Therefore the sui generis right may protect databases on the several levels of abstraction. However, we can still exclude some of the levels. The database sui generis right does not protect physical medium. If an appliance that is used to store a database is stolen, the thief does not infringe the database right, but violates the ownership. How about data? Does the database right protect the bits or the representation in a binary form? No, it does not. Suppose one has a protected database in an IBM mainframe system. If the database is transferred into a UNIX or an MS Windows system, it is possible that the binary representation needs to be changed. For example, letter 'x' in IBM's EBCDIC code is represented as the binary string '10100111' while in ASCII code, 'x' is '01111000'. The indices are probably regenerated. It is possible that most bits are changed while transferring a database from one system to another. Yet, the database right remains untouched – both the original database and the transferred version are protected alike, or from the database right view point, the two databases are the same. Therefore, the database right is not related to data either. Instead, it is related to information. That is, the contents of a protected database need to have some meaning. The large investments that are needed to obtain, to verify, and to present the information are only protected if they are put into results that create value to the users.

The sui generis right provides the maker of a database with the *right to prevent extraction and/or re-utilization* of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. [4] Although individual data items in a database are not protected by the sui generis right, not only the database as a whole is protected but also a substantial part thereof. What is a substantial part? As suggested above, the most important qualification for the sui generis right is significant investment. Therefore, to judge what is substantial, the amount of investment should again be considered. If the investment needed to make a part of a database is significant, that is, if the part alone could be considered to gain the sui generis protection in case it were a separate database, then the part is a substantial part and its extraction and re-utilization without consent is prohibited.

So, the sui generis database right requires *substantial investment*. The investment must be in obtaining, verifying or presenting the contents of the database. If the investment is aimed at something else, it does not constitute the database right. This is illustrated by spin-off doctrine that is especially popular in some courts in the Netherlands. For example, a television program listing, a real-estate listing, and a headlines listing were not databases according to Dutch courts, but merely spin-off products of other activities. On the other hand, Dutch courts have several times held that telephone catalogues and subscriber data are databases. [6] The logic here is not quite clear: it seems that telephone catalogues and subscriber data do require investments, but they are mainly outcome of other activities, namely marketing, customer recruitment, customer service, and the necessary information collection. How large a part of the investment is accomplished just for the catalogues? Probably usually quite small although it is obviously possible to develop a database on subscriber information that needs a lot of investments.

4. PEER-TO-PEER MOBILE DATABASES

A mobile network, in this paper, refers to the computer network to which the end-users connect largely using mobile, wireless appliances. *The Mobile Internet* especially refers to the future mobile network that is relatively open, mostly based on IP protocol, and accessed through many kinds of devices. The sense of *mobility* depends on one's viewpoint. On a protocol level, a significant property of mobility is that the access point is not fixed. This perspective does not necessarily imply that the terminal should be wireless or portable. [9]

On the service level, however, the word *mobile* refers to users' ability to move. Therefore, to be mobile in practice, terminal devices must be wireless and portable. Our focus in this paper is mainly related to the service level. For that reason, we emphasize the wireless and portable properties of terminal devices. [9]

Peer-to-peer or *P2P* is a type of transient Internet network that allows a group of computer users with the same networking program to connect with each other and directly access data and resources in one another's computers. Users can for example share files or spare computer cycles, which makes network a huge distributed computer. [1], [2], [9]

What would a peer-to-peer mobile database be like? Imagine users moving around and accessing a certain service through their mobile devices. The service is not physically located in a central server, but distributed in the user devices on the network. A conventional way to build a service on the Internet is to have a central database, application software on top of the database and users accessing the application through the network. In a P2P solution, both the database and the application is split and distributed into the user devices. That is, each user device includes software that not only provides the user with a view to the service but also shares information in the device with other users. A portion of the database is stored in each device. For performance, quality of service (QoS), and safety reasons, some data are replicated in many devices. No device however needs to have all the data. In other words, the database is not located in one physical place or device, but distributed in a number of mobile devices.

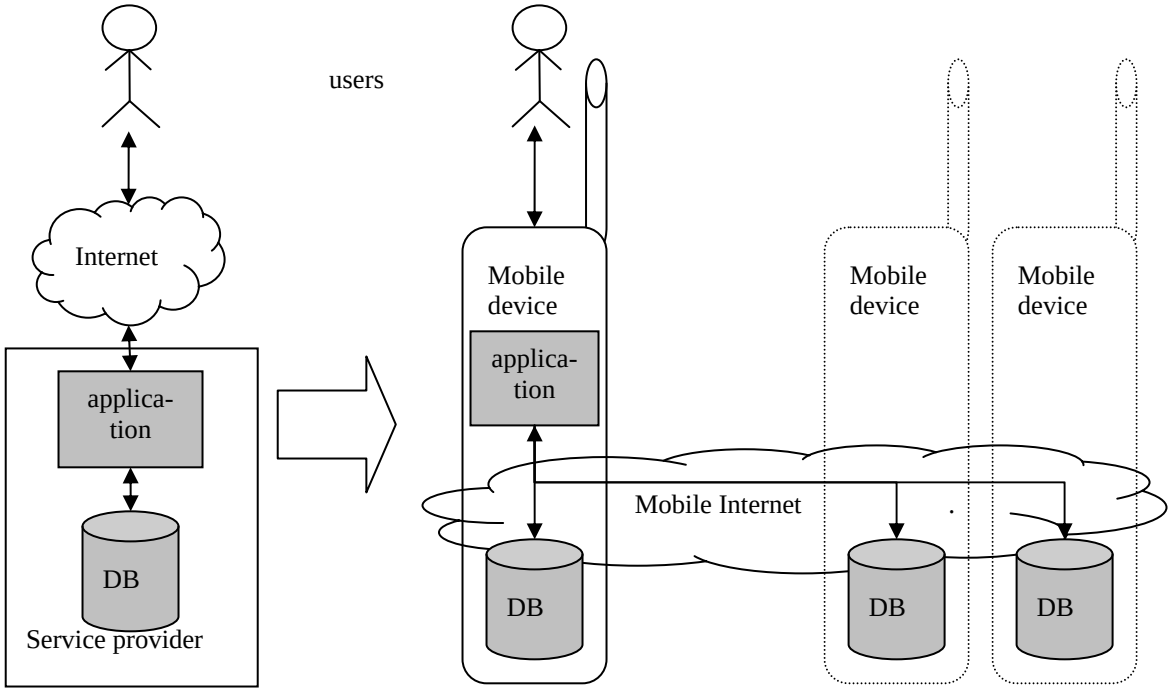

Figure 4. Change from the conventional Internet service model to a mobile P2P model.

Depending on the service, users themselves can produce information and store it in the system, or the system can be used merely to distribute information from other sources to the users. At least some of the data is produced automatically within the system. The application software sends queries through the network to other devices and combines the answers to form a single view to the database. [1], [2] The user does not need to know where the information is located and where the answers come from. To a user, the system at its best will look like whole the service and all the information is in the user's terminal device. This will enable ubiquitous services in which users have always access to enormous databases.

Possibly in the future, P2P systems will be combined with commercial services that add value to the free P2P systems. For example, a free mobile P2P database could be enhanced with commercial value-adding services like updated news or access to some IPR protected resources. Also, companies are implementing digital rights management (DRM) systems into P2P networks making it possible to control the usage of any data in the system.

Obviously, the P2P model has some significant technical preconditions that are not fulfilled so far. For example, the current mobile devices do not have enough computing power and storage capacity nor do mobile networks at the moment provide enough bandwidth to enable this kind of solutions. However, taking the fast pace of technological development into consideration, it seems inevitable that in a few years it will be possible to build such systems.

5. LEGAL IMPLICATIONS

More than ever, it will be troublesome to characterize some databases as arranged in a systematic or methodical way. The physical structure of a database will be in continuous change as the devices move around, access points

change, and connections and routings vary. The momentary snapshot of a database can appear arranged, but after a split second, the arrangement is completely different. Of course, that again depends on the level of abstraction. Certain levels, logical dependencies within database schema, and so on remain unchanged, although devices move. Yet, as discussed above, it is quite unclear on what level of abstraction database right requires certain arrangement.

We have concluded above that *the requirement of substantial investment is central in the database sui generis right*. Will the mobility or peer-to-peer approach change something in investments? In general, the mobility will be achieved with the help of enabling infrastructure and middleware. Those who build databases will not usually need to worry much about technical details related to mobility and peer-to-peer approach. Although significant investments will be required to develop sophisticated technologies to enable mobile P2P databases, they will not be investments in a particular database and they do not help to achieve the sui generis right. Instead, investments in a database as such will not change much. Also in the mobile P2P databases, there will be qualitatively and quantitatively substantial investments in the obtaining, verification and presentation of the contents. However, if the users will obtain, verify, or present the contents themselves in a peer-to-peer fashion, then it is likely that no single person or entity has contributed substantial investments. Such a P2P database may remain outside of sui generis right. That, nevertheless, is probably desirable. Most users, in all likelihood, prefer that no-one gets exclusive rights in the outcome of their joint effort. Another interpretation could be that if the total investment is substantial, then the database is protected and all the users that have contributed get a collective right. In practice that kind of collective right is very difficult to manage and does not necessarily satisfy users' expectations. The directive nonetheless does not tell us which interpretation is correct. Yet, if peer-to-peer technologies are used only to deliver a database to users, but the content is obtained, verified, and presented by a single entity, a service provider, then this entity will have the database right.

The Mobile Internet will be significantly international. It means that mobile P2P databases can spread among different countries effortlessly. The rights in databases nonetheless depend heavily on jurisdiction. Within European Union the database sui generis right brings forth a common legal ground for business models based on mobile databases. However, as a mobile database spreads further, the legal situation becomes more complex. From international perspective it would be desirable for the mobile business and ubiquitous services that countries adopt similar database protection laws.

In practice, however, efficient DRM systems may solve many of the legal uncertainties. An efficient DRM is also able to manage database rights to the information. From business perspective, therefore, it is a sound strategy to implement a DRM system in order to control the usage of databases. DRM can protect all layers of data regardless to its semantic characterization as information or data, and regardless to its representation and value to the user.

To conclude, the definition of 'database' in the directive is questionable. It is hard to tell on which level of abstraction it refers to. Instead, the requirement of significant investment is very important. Parallel to originality, novelty and non-obviousness, distinctiveness, and so on, it adds a vital new area of subject matter into the field of intellectual property.

In general, the present arguments in favor and against the database sui generis right will stay with respect to the future mobile peer-to-peer databases. Some of the problems that are already visible will be highlighted. New technical solutions may solve some of the problems in practice. However, the fundamental idea behind the database right, that of protecting substantial investments, will remain central as regards to mobile P2P databases.

6. ACKNOWLEDGEMENTS

The Finnish National Technology Agency *Tekes* as well as our industrial partners have generously supported our work. Discussions with our colleagues Dr. Jukka Kemppinen, Dr. Martti Mäntylä, Mr. Ville Oksanen, Mr. Tommo Reti, Ms. Aura Soininen, Mr. Risto Sarvas, Mr. Yki Kortesniemi, and others have been most instructive.

REFERENCES

[1] K. Aberer, M. Punceva, M. Hauswirth, R. Schmidt. Improving Data Access in P2P Systems. IEEE Internet Computing Vol. 6, no 1, January/February 2002.

[2] D. Clark. *Face-to-Face with Peer-to-Peer Networking*. Computer Vol. 34, no. 1, January 2001.

[3] Copyright and Rights in Databases Regulations 1997, Statutory Instrument 1997 No. 3032, the United Kingdom, 1997.

[4] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, 1996.

[5] P.-L. Haarmann. *Tekijänoikeus & lähioikeudet*. (In Finnish) Kauppakaari Oyj, !999.

[6] P. B. Hugenholtz. *The New Database Right: Early Case Law from Europe*. Ninth Annual Conference on International IP Law & Policy, Fordham University School of Law, New York, USA, 2001

[7] H. Korth, A. Silberschatz. *Database System Concepts*. 2nd Ed. McGraw-Hill. 1991.

[8] R. P. Merges, P. S. Menell, M. A. Lemley. *Intellectual Property in the New Technological Age*. 2nd Ed., Aspen Law & Business, 2000.

[9] O. Pitkänen, M. Välimäki, J. Kemppinen, M. Mäntylä, Assessing Legal Challenges on the Mobile Web, *Proceedings of M-Business-2002*, Athens, Greek, 2002.

[10] Tekijänoikeuslaki (Finnish Copyright Act) 404/1961, last changed 748/1998, 1999.

[11] Whatis.com, http://whatis.techtarget.com/