DATA MINING, SUI GENERIS RIGHT AND AGGREGATORS. SOME REFLECTIONS AFTER THE JUDGMENT OF THE COURT OF JUSTICE OF THE EUROPEAN UNION IN CASE CV-ONLINE LATVIA V MELONS

Gemma Minero Alejandre*
Universidad Autónoma de Madrid

Abstract

European Intellectual Property Law provides a special intellectual property (IP) right –the sui generis right– for those database makers that made a substantial investment when creating the database. This IP right is infringed in those cases where the defendant creates a risk that the database maker will lose incomes, thereby depriving that database maker of revenue which should have enabled him to redeem the cost of the investment in setting up and operating the database. However, in recent Melons case, the Court of Justice of the European Union redefines the sui generis right in relation to online aggregators. It highlights the fact that access to information and competition concerns have to enter the sui generis right's infringement test. It is the first time that this Court balances the interest of the database maker and those of the other parties, as part of the infringement test. We have to keep in mind European caselaw on aggregators and scraping since scraping tools are playing and will play a key role in the development of artificial intelligence.

Keywords: data mining, databases, aggregators, investment.

MINERÍA DE DATOS, DERECHO SUI GENERIS Y AGREGADORES. ALGUNAS REFLEXIONES TRAS LA SENTENCIA DEL TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA EN EL CASO CV-ONLINE LATVIA V MELONS

Resumen

En el Derecho europeo de la Propiedad intelectual se regula un derecho especial –el llamado derecho sui generis– para aquellos fabricantes de bases de datos que llevaron a cabo una inversión sustancial en la creación de esa base de datos. Este derecho de propiedad intelectual se ve infringido en aquellos supuestos en los que el demandado priva al fabricante de los ingresos que le permitirían amortizar el coste de la inversión hecha en la constitución de la base de datos y en su funcionamiento. Sin embargo, el Tribunal de Justicia, en una reciente sentencia en el caso Melons, ha redefinido el derecho sui generis en relación con los agregadores en línea. Considera que tanto el acceso a la información como los intereses de sujetos competidores han de ser parámetros a tener en cuenta cuando se analiza si ha existido o no una infracción del derecho sui generis. Es la primera vez que introduce entre los parámetros del análisis de la infracción la necesidad de ponderar los intereses del fabricante de la base de datos y los intereses de terceros. Este pronunciamiento habrá de tenerse muy presente, toda vez que los agregadores y las herramientas de *scraping* jugarán un papel clave en el desarrollo de proyectos de inteligencia artificial.

Palabras clave: minería de datos, bases de datos, agregadores, inversión.

DOI: https://doi.org/10.25145/j.anfade.2022.39.04
Anales de la Facultad de Derecho, 39; julio 2022, pp. 63-72; ISSN: e-2530-8319

1. CONTEXT OF DATABASES PRODUCTION AND ARTIFICIAL INTELLIGENCE DEVELOPMENT AMONG THE EUROPEAN UNION

In the EU context we have a special –and only European– IP right –the sui generis right– for those database makers that made a substantial investment when creating the database, whether it contains works or not, and whether it contains personal or non-personal data.[1]

That substantial investment can be made by just one person, but, in many cases, it is the result of the activities of many people and/or some undertakings processing and aggregating data. In some cases, the database created is the result of a data mining process.

On the other hand, we all have in mind that data mining technologies are fundamental for the ongoing of our current digital society. This is the reason why the legal framework of data mining has become a current topic. The Single Market Directive contains a specific regulation of two kinds of data mining exceptions to the sui generis right in Articles 3 and 4.[2] One for research purpose and a second one for other aims. However, this second one is not a real exception as such since the right holder can decide to opt out. Exceptions to IP rights, by definition, are mandatory for the right holders.

In the modern digital economy, data are being considered the 'new oil' and the sui generis right might be used to control any access to the database, thus having an undeniable relevance. That is why current legal debate about data economy as well as databases protection discussions get increasing attention. In particular, this paper is focused on the database protection debate.

---

* Profesora contratada Doctora de Derecho Civil de la Universidad Autónoma de Madrid.

[1] This right was harmonized by Database Directive (Directive 96/9/EC of the European Parliament and of the Council, of 1 March 1996, on the legal protection of databases).

[2] Directive 2019/790 of the European Parliament and of the Council, of 17 April 2019, on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. Exceptions for data mining contained within articles 3 and 4 of this Directive have been studied by Geiger, CH., Frosio, G. and Bulayenko, O., "Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU" in Saiz García, C. and Evangelio Llorca, R. (dir.) *Propiedad intelectual y mercado único digital europeo,* Tirant Lo Blanch, 2020, pp. 36-45; Jiménez Serranía, V., "Datos, minería e innovación: ¿Qvo Vadis, Europa? Análisis sobre las nuevas excepciones para la minería de textos y datos," *Cuadernos de Derecho Transnacional*, 2020, vol. 12, n. 1; Sánchez Aristi, R. y Oyarzabal Oyonarte, N. "Decadencia y caída del Texto Refundido de la Ley de Propiedad Intelectual: la transposición de la Directiva 2019/790 sobre derechos de autor en el mercado único digital por el Real Decreto-Ley 24/2021, de 2 de noviembre," *Pe.i. Revista de Propiedad Intelectual*, N-69, September-December 2021; Hilty, R. and Richter,H., "Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3-Text and Data Mining)," 14 January 2017. *Max Planck Institute for Innovation & Competition Research Paper*, n. 17-02; and González Otero, B., "Las excepciones de minería de textos y datos más allá de los derechos de autor: la ordenación privada contraataca," in Saiz García, C. and Evangelio Llorca, R. (dir.) *Propiedad intelectual y mercado único digital europeo*, Tirant-lo-Blanch, 2020, p. 98.

The explanation is complex, since databases protection and data mining are like a snake eating its own tail. If an artificial intelligence project uses a database that is already protected by the sui generis right, the rightsholders' authorization is needed. However, in some cases, data mining exceptions apply, and that authorization will not be needed. In any case, if the company managing that artificial intelligence project makes data mining, the result of that process could be the creation of a new database, in order to create trends and correlations, and that new database could be also eligible for new sui generis right, provided a substantial investment was made. Taking all these elements into account, a first conclusion is that databases production is a kind of eternal process under construction. In our economies, different actors produce data, some others collect them and create databases and some others analyse them using algorithms thus enabling new insights (and new databases).

According to definition contained within Article 2 of Digital Single Market Directive, data mining means any "automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations." Text data mining turns texts (or images or audio-video) into data. It is essential for any kind of search engine; it is essential for any kind of statistical analysis of large bodies of content; and it underpins many machine learning applications. It is fundamental for feeding AI initiatives. They all need large amounts of data. Machines cannot "learn" unless the things they want to learn about have already been rendered into data, and that involves data mining.

Not only machines, but also humans need data mining. No researcher can read all relevant research papers that are published in his field of interest. Even if he or she could, he would not be able to find patterns in the research results that emerge only from data mining. Data mining makes the analysis of vast amounts of information possible, that would not otherwise be possible. Through data mining, researchers can discover new knowledge from existing knowledge. Therefore, it is an important research tool.

That makes the EU a strategic place to invest, a strategic place to create a company if we want to develop an artificial intelligence project that requires the use of data mining technologies and that implies the creation of a second database.

2. AN APPROACH TO THE EUROPEAN PROTECTION OF DATABASES

The so-called sui generis right is a special European intellectual property right on databases, which was harmonized in 1996. It is currently one of the main tools for databases protection.

The rightholder of sui generis right is the person who takes the initiative and the risk of investing. Considering that this special protection provides no moral rights to the database maker –such as the paternity right or the integrity right of authors–, the European legislator decided to recognize this protection both to natural persons and companies that bear the risk of the investment, provided they have a direct link

with the European territory. In the case of natural persons, the nationality of an EU Member State or the residence in the EU territory are required. In relation to undertakings, it is required that the company was formed in accordance with the law of an EU Member State and having the registered office, central administration or principal place of business within the EU.[3]

According to Article 7 of the Database Directive, sui generis right arises when there has been a substantial investment in either the obtaining, verification or presentation of the contents of the database. It is irrelevant whether the contents are already protected by copyright or by a related IP right or not. The right holder has the right to decide whether to authorize or not extractions of contents and reutilizations and to decide the conditions, such as time and price. However, this right holder has no right to impede the creation of new comparable or competitor databases containing the same or similar contents in those cases where the second maker does not use his/her database as the source of the new dataset.

Artificial intelligence databases and databases containing machine-generated data are eligible for the sui generis right, provided there is a substantial investment. This conclusion follows the logic of the non-authorship or no human characterization allowed by the Database Directive. Copyright protection requires human authorship. On the contrary, sui generis protection just requires investment, so that it is irrelevant whether the activities of obtaining, verification or presentation of the contents are done by a natural person or by an algorithm.

The substantial nature of the investment could be quantitative or qualitative, either in resources, efforts, or time. It has to be analysed in a case-by-case decision, since the Database Directive does not provide a concrete minimum quantity of money. There is no doubt that many current artificial intelligence tools used for the data processing imply substantial investment eligible for the sui generis protection.

The Court of Justice of the European Union has narrowed the scope of sui generis right by considering that investment in the creation ex novo of data or contents of the database is not protected by the sui generis right. The database maker has to invest resources on processing data that already existed, that is, resources in pre-existing data, since those resources made on the preliminary phase of generation of data are excluded from protection.

The Court of Justice of the European Union has highlighted this idea in the *Fixture Marketing saga* and in the *Football Dataco saga* that were handed down in 2004 and in 2012. The company creating data, such as sport fixtures, does not own sui generis right because the investment made was only referred to the creation of data.[4] In those cases, the database maker will only be protected by the sui generis right provided there was an additional investment in the resources used for the

[3] The sui generis right is currently just a European special IP right. There is no equivalent IP protection in other countries around the world. It has not been harmonized worldwide in any international treaty and it is no longer in the World Intellectual Property Organization's agenda.

[4] Judgments of the Court of Justice of the European Union of 9 November 2004, Fixtures Marketing Ltd v. Oy Veikkaus Ab; Fixtures Marketing Ltd v. Svenska Spel AB; and Fixtures Marketing

verification or presentation of the contents or for processing contents that already exist, for example, using some data mining tools.[5] Investment in data mining tools for the creation of new contents is also excluded from protection.[6]

The right holder of the sui generis right has to be aware that he or she is part of a data value chain, so other stakeholders of that chain will be interested on buying a licence of that database in order to use the contents together with other data processing tools, such as their own algorithms, for the creation of new enriched data and a new database containing them. Provided the new database implies a new substantial investment, the licensees will have new and independent sui generis right. The fee that was paid for the licence to use the contents of the first database will be eligible for the sui generis protection provided it is a substantial investment and it does not refer to the creation of contents, but to the processing of contents that already exist.

When analysing potential sui generis right infringements, good or bad faith of the users is irrelevant. The traditional logic of copyright is also applicable to the sui generis right. Consequently, the right holder can sue any unlawful user –any person that has no right to access to the database– or any lawful user that used the database in a manner or for a purpose that goes beyond the authorization or licence given by the right holder in such a way that it prejudices the right holder's legitimate interests or conflicts with normal exploitation of the database, that is, in such a way that deprives the database maker of revenue which should have enabled him/her to redeem the cost of the investment in setting up and operating the database.

According to Article 10 of the Database Directive, the term of protection of the sui generis right is only fifteen years. Such a short term of protection is

---

Ltd v. OPAP; and Judgment of the Court of Justice of the European Union of 1 March 2012, Football Dataco Ltd and Others v. Yahoo! UK Ltd and Others.

[5] That additional substantial investment was satisfied in the facts of the judgment of the Court of Justice of the European Union 18 Oct. 2012, Football Dataco v. Sportradar. The claimants alleged infringement by the defendants of the sui generis right, which the claimants claimed to have in a database relating to football matches in progress. That database was a product different to just football fixtures. It was a compilation of data about football matches in progress (goals and scorers, yellow and red cards and which players were given them and when, penalties and substitutions, etc.). The data was said not to be created, but just collected, mainly by ex-professional footballers who worked for Football Dataco and attended the matches for this purpose.

[6] Derclaye and Husovec have highlighted the fact that the notion of substantial investment is one of the most problematic provisions of the Database Directive. "Two databases of the same kind created by two companies can lead to entirely diverging outcomes. One might be protected as a result of the relevant investments, while the other might not. For instance, a dataset of flight connections created by airlines selling the flights, will not enjoy protection. However, purchasing the same data set could qualify as relevant investment." See Derclaye, E. and Husovec, M., "Sui Generis Database Protection 2.0: Judicial and Legislative Reforms," *European Intellectual Property Review*, 2022, 44(6), p. 325. It is noteworthy that the Evaluation Report made by the European Commission in 2018 concludes that the exclusion of investment related to the creation of contents of the database should also apply to machine generated databases and databases automatically generated by Internet of Things devices.

remarkable if we compare it with other IP rights. However, any new substantial investment will be eligible for new independent sui generis right, with its own term of protection, that is, for another fifteen years from the date of the new investment. This potential nearly eternal renewal possibility is controversial. In particular, in sole source data situations.

3. THE JUDGMENT OF THE COURT OF JUSTICE OF THE EUROPEAN UNION IN MELONS CASE

It is not the first time that the Court of Justice of the European Union talks about web scraping and aggregators and sui generis right, but it is the first time that it says what it has said. We have to keep in mind European caselaw on scraping since scraping tools are playing and will play a key role in the development of artificial intelligence.

The question of the application of sui generis right to results provided by metasearch engines was raised before the Court of Justice of the European Union in *Innoweb v. Wegner* case in 2013.[7] The Court gave an affirmative answer: the sui generis right may be used to protect the investment made in a website which provides access to an online collection of car advertisements and contains an internal search engine. According to the ECJ, the right holder of the sui generis right can implement technical measures to block bots commonly used for web-scraping, since web-scraping implies extraction and reutilization of the contents of the database and, therefore, the right holder of the sui generis right can decide not to authorize these uses, and to avoid them with technological protection measures.

The defendant was the operator of a dedicated meta-search engine that makes it possible to search the entire contents of a database in real time, by entering an end user's query in the search engine of the database. Therefore, the defendant uses the plaintiff's internal search engine.

The judgment found against the meta-search engine, even though it displayed hyperlinks to the contents of the plaintiff's website. According to the Court of Justice of the European Union, the defendant came close to becoming a 'parasitic competing product', since many end users no longer have any need to go to the plaintiff's website. Consequently, it creates a risk that the database maker will lose income, in particular the income from advertising on his website, thereby depriving that database maker of revenue which should have enabled him to redeem the cost of the investment in setting up and operating the database. According to the Court of Justice, the investment made by the database maker will be affected, since the number of visits to his website will decrease and, consequently, the publicity

[7] Judgment of the Court of Justice of the European Union of 19 December 2013, Innoweb BV v Wegener ICT Media BV and Wegener Mediaventions BV.

revenues will also decrease.[8] Taking all these elements into consideration, the Court of Justice concluded that the act of making available on the Internet a dedicated meta search engine re-utilises the whole or a substantial part of the contents of a database protected by the sui generis right. Therefore, it infringes the sui generis right.

On the contrary, in recent *Melons* case, the Court of Justice of the European Union redefines the sui generis right.[9] It highlights the fact that access to information and competition concerns have to enter the sui generis right's infringement test.

In a nutshell, the facts of the case were the following: a specialist search engine engaging in re-use of substantial parts of the database of a job adverts website was accused of violating sui generis database right.[10] Unlike in Innoweb case, in Melons case, the defendant –this specialist search engine or aggregator– does not use the search function of the plaintiff's website but develops its own way to explore the database. Moreover, the user is only offered deep links.

According to the Court of Justice of the European Union, the defendant extracts and reutilises the database. It implies previous copies and indexes of the content of the plaintiff's database. However, the Court of Justice of the European Union says that an extraction or reutilization of the contents is not enough for concluding that there was an infringement. It only constitutes sui generis right infringement provided such uses deprive the database maker or endanger its right to redeem the cost of his investment. This important aspect introduces the leitmotiv of the database protection into the infringement test. For the Court, the infringing must imply "significant detriment" to the database maker's investment. If there is not such a detriment, then there is no infringement. The Court of Justice does not say it directly, but it can be inferred that in the case the defendant creates a new product, the risk for the plaintiff that his or her investment may not be redeemed must be high or evident, not just potential and the plaintiff must prove it.

Moreover, while analysing whether there is a significant detriment to the investment, national courts must balance the interest of the database maker and of the other parties, as part of the infringement test. The court explicitly mentions the legitimate interests of "users" to have access to information contained in the database

[8] In this case, the Court of Justice pointed out that the defendant's activity undermines the sui generis right, since the user no longer has any need to proceed via the homepage –the plaintiff's homepage–, since he can explore that database directly using the service of the operator of the meta search engine –the defendant's service–.

[9] Judgment of the Court of Justice of the European Union of 3 June 2021, case CV-Online Latvia v Melons. About the alteration of the trajectory of sui generis right that this recent judgment introduces, see Derclaye, E. and Husovec, M., "Sui Generis Database Protection 2.0: Judicial and Legislative Reforms," *opus cit*., pp. 323-331. According to these authors, the Court of Justice gives a more measured and mature view of the sui generis right.

[10] A search engine such as that at issue in the proceedings makes it possible to explore all the contents of the website and provides its users access to the entirety of the contents, but using means other than that provided for by the maker of the first database.

and "competitors" to create innovative products based on that information.[11] The Court notes that aggregators contribute to the creation and distribution of new and better products and services with added value in the information sector. By offering their users a unified interface enabling them to search several databases, they allow the information on the internet to be better structured and to be searched more efficiently. They also contribute to the smooth functioning of competition and to the transparency of offers and prices. Finally, the Court notes that it is up to the national court to apply competition law to the case at hand.[12] We must keep an eye on whether the Court of Justice of the European Union confirms this doctrine.

As for the future, taking into account Melons doctrine, it could be noted that very innovative products producing strong consumer benefits or socially important re-use of data that do not have significant impact on the investments made are likely to prevail. However, in other not so innovative cases where there is a high-risk for database makers and low gain for users, could be considered parasitic or close to parasitic products.

4. SOME REFLECTIONS ABOUT NEW DATA MINING EXCEPTIONS

As it was stated before, the Digital Single Market Directive regulates two different exceptions for data mining: one specific scientific research exception and a general data mining exception. The European legislator is aware of the leading role of data mining technologies in current digital economy and, in particular, of its benefits for the research community.

In particular, Recital 8 of the Digital Single Market Directive mentions universities and other research organisations and cultural heritage institutions, such as archives, libraries and museums, as the main beneficiaries of the first data mining exception, that is, the exception contained within Article 3 of this Directive. I am referring to data mining for scientific research exception.

According to Recital 12, these beneficiaries have in common that they act either on a not-for-profit basis or in the context of a public-interest mission. Such a

[11] The Court of Justice has introduced the fair balance parameter, so that the sui generis right has to be balanced with other rights. Contrary to recitals 4 and 9 of Directive 2001/29, of 22 May 2001, on the harmonisation of certain aspects of copyright and related rights in the information society, the Database Directive does not state that the protection of databases must be high.

[12] We cannot forget that Article 13 of the Database Directive states the provisions of this Directive are without prejudice to the competition rules of EU law or that of the Member States. This conclusion of the Court of Justice has been strongly criticized by Derclaye and Husovec, which state: "It should be clarified that sui generis database protection pre-empts national slavish imitation/parasitism." "Such a provision would make it clear that it is impossible to cumulate slavish imitation or parasitism with the sui generis right, or even extend it beyond the EU Law." See Derclaye, E. and Husovec, M., "Sui Generis Database Protection 2.0: Judicial and Legislative Reforms," *opus cit.*, p. 330.

public-interest mission could, for example, be reflected through public funding. For example, a public hospital using data mining for optimizing medical treatments and organizing beds, surgery rooms and other spaces. On the contrary, organisations upon which commercial undertakings have a decisive influence are excluded. For example, insurance companies using data mining for finding abuses and frauds.

However, recital 11 also states: in line with private sector's collaboration policies, research and cultural organisations also benefit from this exception when their "research activities are carried out in the framework of public-private partnerships." Therefore, research organizations can also apply this exception when using private technological tools for data mining.

In any case, the Directive requires lawful access. It covers access to content based on an open access policy, or through contractual arrangements between rightholders and the beneficiaries of this exception, such as subscriptions.

Taking into account that this exception is limited to entities carrying out scientific research, any potential harm created to right holders would be minimal. This is the reason why the Directive does not require compensation.

Apart from this mandatory exception contained in article 3 and referred to scientific research, the Directive also regulates a second data mining exception in Article 4. Therefore, by definition, this second kind of data mining applies to the private sector or to public entities, but for aims different to scientific research, such as developing governmental services.

Actually, it is not a real exception as such, but a reminder for rightsholders. This optional data mining exception applies on condition that the use of the database has not been expressly reserved by their rightsholders. For example, rightsholders can decide to do this reservation in the conditions of the website or in the text of the contractual agreements and, of course, rightsholders can apply technological measures to ensure this reservation. Consequently, the right holder of the sui generis right on a database can decide to reserve the extraction right for data mining or he can also decide to allow data mining of the whole or of a part of his or her database, in exchange of a price or under other conditions. The right holder is not obliged to allow data mining of his or her database to any kind of users for any kind of aim.

To sum up the European approach, there is not a broad exception for data mining for purposes others than research that could play a fostering role for artificial intelligence development among the European countries. That despite the fact that some artificial intelligence techniques work in a similar way to aggregators. Besides, we cannot forget that any exception and limitation must be interpreted narrowly.[13]

[13] WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment, prepared by S. Ricketson, April 2003, https://www.wipo.int/edocs/mdocs/copyright/en/sccr_9/sccr_9_7.doc.

5. SOME CONCLUSIONS

In a recent judgment, the Court of Justice of the European Union has redefined the sui generis right, in order to get a balance between IP rights and Competition Law. It is the first time that the Court of Justice does it, so we still have to wait and see whether it continues this interpretation or, on the contrary, whether it comes back to its previous doctrine. There is no doubt that with this ruling the Court of Justice has narrowed the sui generis right.

This result might have a double effect. On the one hand, this interpretation might be used to foster artificial intelligence, whose creation needs previous data mining activities. On the hand, we cannot deny that the European sui generis right is also a reason fostering artificial intelligence, since the datasets that are created when doing data mining might be eligible for the sui generis protection among the European countries.

In those cases where the Court of Justice has understood that there is no infringement of the sui generis right, then there is no reason to study whether the new data mining exceptions apply or not. If there is no infringement, the use is legal and, therefore, there is no need to apply any exception.

Moreover, we cannot forget one important detail: the database in dispute in *Melons* case did not contain works or other subject-matter. It contains mere data or information. Maybe the answer would not be the same if the defendant would have used a database containing works for the creation of a second database. In that case, it would be less probable that the Court of Justice states that it is necessary to strike a balance between, on the one hand, the legitimate interest of the makers of the database in being able to redeem their substantial investment and, on the other hand, that of competitors in having access to works contained in the previous database and the possibility of creating new products based on that contents. In other words, in this hypothetical case, the resulting product would not contribute to the development of the information market. Let's see if the Court of Justice agrees with us or not...

Recibido: 01/09/2022; aceptado: 20/09/2022