

# Database protection

Caterina Sganga

Oftentimes, the value of data greatly increases when they are “cleaned”, verified, organized, linked to each other and presented in a manner that makes them easily searchable, readable and manageable. (Seuba-Geiger-Penin 2018, 85-94; EC Data Market Study 2017, 5). The process of collecting, verifying, classifying and structuring data, however, may require substantial investment of money and work. In some cases, the resulting compilation presents original features as to the structure and organization of its content. In other instances, originality is not present, but the realization of the database has still entailed high production costs – a circumstance that increases the risk of free riding, which negatively impacts on the possibility for the producer to get an adequate return on investment. (Database Directive Explanatory Memo 1996, 2)

The question of whether, why and how to protect databases started engaging national legislators much before the advent of big data and artificial intelligence (AI) agents. Judicial decisions on the matter trace back to the 1970s, while the main legislative efforts developed in the 1990s (Derclaye 2008, 11 et seq.). In an attempt to adapt old provisions on anthologies and other collections of works to cover compilations of data, despite the non-copyrightability of the latter, international sources have set a minimum standard of protection for databases. At the same time, however, national systems have started diverging on, *inter alia*, subject-matter of protection, standard of originality required, rights conferred, aligning around two basic models which can be ascribed respectively to the European Union and the United States. The lowest the protection offered via copyright law or other *sui generis* rights, the more frequent the recourse is to trade secret or alternative forms of protection of structured data corpora.

Due to the extreme relevance of database protection for data science, the following pages will offer an overview of the international regulation in the field, and a comparison of the two main models of national/regional protection, concluding with an analysis of the exceptions and limitations to database copyright and *sui generis* right (on the protection offered by trade secret law, see entry *Trade secret*).

## 1. International sources

The first international treaty intervening to harmonize basic principles and definitions in the field of copyright is the Berne Convention (BC, 1886), revised multiple times until 1971 (Ricketson-Ginsburg 2006). Each revision added new substantive provisions, among which Article 2(5) BC, which provides protection for “collections of literary and artistic works such as encyclopedias and anthologies which, by reasons of the selection or arrangement of their contents, constitute intellectual creations”. The references to “literary and artistic works” suggests that the definition is confined to collections of preexisting works, with the exclusion of compilations of non-copyrightable materials such as data (Goldstein and Hugenholtz 2010, 206). However, interpretations of the text diverge. For instance, in 1982 the WIPO-UNESCO Committee of Governmental Experts maintained that the text of the Berne Convention offers protection against the reproduction of databases which consist of uncopyrightable data, arguing that “collections and compilations of information” could qualify for copyright protection (Second Committee 1982, 245).

Article 2(5) BC sets a relatively low standard of originality, requiring the selection and arrangement to constitute an “intellectual creation”. According to the WIPO Guide to the Berne Convention, the provision expects the maker to “bring to bear an element of creativity; merely listing the works or extracts without offering any personal contribution is not enough” (Guide to Berne 1978, 20). The standard, as we will see below, has been interpreted in a largely different manner by Member States.

The first mention to databases as a subject-matter of copyright can be found in the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs, 1994), which created a new minimum standard of protection for intellectual property rights for all the states members of the WTO (Gervais 2017). Article 10(2) TRIPs states that “compilations of data or other material, which in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such”. The specification is particularly useful to mark a distinction with the “collection of works” mentioned by the Berne Convention, which the TRIPs makes binding for WTO Member States by requiring compliance to Articles 1 to 21 BC (Article 9 TRIPs). Article 10(2) TRIPs clarifies that the protection demanded by the Treaty does not extend to the data or material constituting the content of the database, and should be without prejudice to any copyright or other exclusive rights subsisting on the latter. The provision implicitly assumes that some category of data may indeed be copyrightable, as long as they do not fall into categories such as “ideas, procedures, methods of operation or mathematical concepts as such”, excluded from the subject-matter of copyright according to Article 9(2) TRIPs.

The WIPO Copyright Treaty (WCT, 1996), enacted as a special agreement under Article 20 BC to overcome the standstill caused by the lack of a majoritarian consensus on the amendments to the Convention, follows the same approach adopted by the TRIPs (Ficsor 2002). Article 1 WCT requires contracting parties to comply with Articles 1 to 21 BC, and specifies that nothing in the Treaty shall derogate from existing obligations or prejudice any rights and obligations under the Convention. More specifically, Article 3 WCT provides that parties should apply *mutatis mutandis* Articles 2 to 6 BC in respect of the protection provided for in the Treaty, thus including also Article 2(5) BC on collection of works. Then, Article 5 WCT protects – “as such” - “compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations”, it excludes from the scope of the right “the data or the material itself”, and specifies that the exclusivity on the database is independent and not in prejudice “to any copyright subsisting in the data or material contained in the compilation.” No preparatory works or Agreed Statements to the WCT clarifies whether the provision should be understood as an expansion of Article 2(5) BC only, or it should be more broadly interpreted as a broadening of the notion of protected work under Article 2 BC as a whole, including also data and their compilation (Goldstein-Hugenholtz 2010, 206). It is uncontested, instead, that both the definitions offered by the WCT and the TRIPs Agreement are wide enough to include also electronic databases.

Slightly after the adoption of the Database Directive (96/9/EC) in 1996, the European Commission submitted a proposal for an international treaty on the *sui generis* protection of databases to the WIPO Committees of Governmental Experts, inspired by its new legislation and aimed at the approval of a new Berne Protocol (WIPO doc BCP/CE/VI/13, 1 February 1996). The proposal was supported by several delegations, slightly amended by the US (WIPO doc BCP/CE/VII/2 – INR/CE/VI/2, 20 May 1996), and finally converged into a consolidated text prepared by the Chairman of the Committee

for the WIPO Diplomatic Conference 1996. The Conference, which discussed and adopted the WCT and the WIPO Performances and Phonograms Treaty (WPPT), did not make it in time to negotiate the proposed Database Treaty, and only adopted a recommendation to schedule the discussion in the first quarter of 1997. After many years of iterations, the interest surrounding the Treaty decreased, until the matter disappeared from the WIPO agenda (Walter-von Lewinski 2010, 22.01-22.04).

## **2. The great divide: the EU copyright / *sui generis* protection vs the US minimalist approach**

On the basis of the input provided by international conventions and by the circulation of legal solutions, most national copyright statutes introduced a protection for compilation or collections of protected works, either by providing a general definition or by introducing in their list of protected subject matters anthologies, encyclopedias and the like (Davison 2003, 11). The major difference between legal systems lays in the definition of the minimum threshold required for a collection to be protected. Countries belonging to the continental tradition have generally been keener to require the presence of intellectual creativity and an individual touch to consider the originality requirement met. On the contrary, the Anglo-Saxon model, and particularly the British system, has generally opted for the “skill, judgment, labour and learning” standard, also known as “sweat of the brow” (see *Macmillan & Co. Ltd v Cooper* [1923] 40 TLR 186, 188 (1924), and *Independent Television Publications Ltd v Time Out Ltd* [Ch.Div.1984] FSR 64), where protection is conferred for the mere effort put in the creation of the collection, with little or no regard to its originality and creativity (for a comparative overview, see Davison 2003; Derclaye 2008).

With the steady development of software and storage tools from the 1980s on, and the consequent increase in value and pervasiveness of the data collection market, the difficult adaptability of old provisions, tailored to compilation of protected works, on collections of non-copyrightable materials became particularly evident and triggered a debate on the need for a legislative reform (Hugenholtz 2000, 183). In the EU, the first document mentioning the need for immediate action in order to maintain the competitiveness of European industries was the Green Paper on Copyright and the Challenges of New Technologies (1988), launching a consultation on the main features of potential harmonization of database protection in the Union. After four years of debate, on 11 March 1996 the Council and Parliament approved the Database Directive, later defined by the Commission as “a cornerstone of intellectual property protection in the new technological environment (Follow-up to the Green Paper on Copyright in the Information Society, 9).

The Directive protects databases in any form – thus both in electronic and non-electronic format (Recital 14 Database). Article 1 defines a database as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”, covering also the materials necessary for the operation or consultation of certain databases (eg thesaurus and indexation system), but with the exclusion of software used in their making or operation. Recital 17 carves out from the scope of protection “a recording of an audiovisual, cinematographic, literary or musical work as such”, while Recital 18 does it with the compilation of several recordings of musical performances on a CD, both because not original enough to meet the conditions for copyright protection, and because it does not entail a substantial enough investment to justify the attribution of a *sui generis* right.

The Directive introduces an innovative two-prong regime of protection. The structure of the database is protected by copyright when the database itself, “by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation” (Article 3). The exclusive rights conferred are that of reproduction, translation, adaptation, arrangement, distribution, communication, display and performance to the public (Article 5). The protection does not extend to the content of the database, which remains protected by the respective exclusive rights if any (Recital 26). While the notion of “selection” refers to specific materials and not to the overall subject matter, the arrangement should be original in the method and system used (Walter-Von Lewinski, 2010, 9.3.5).

Commentators have long debated on the criteria to determine the originality of arrangements in electronic databases, identifying them in the way how the content is presented on screen on the basis of the command structure (Pattison 1992, 116; Beutler 1996, 325; Dreier 1992, 746), in the design of a data model in case of database of factual information (Wiebe 2004, 71), in the access and retrieval system (Walter-Von Lewinski, 2010 9.3.6), and so forth. More generally, national courts have not fully aligned on the degree of originality required for a database to be protected, due to the different standards adopted by Member States in their copyright case law (Rosati 2013, 64 et seq.; Davison 2003, 13-17). To avoid short-circuits, Article 14(2) Database has introduced a transitional provision, leaving unaffected the copyright protection of existing databases not fulfilling the criteria of protection set by Article 3(1) Database. However, the CJEU has already intervened on the matter, harmonizing the notion of originality in EU copyright law (from case C-5/08, *Infopaq International A/S v Danske Dagblades Forening*, EU:C:2009:465) and ruling specifically in the field of databases that the test of “the author’s own intellectual creation” requires that the selection or arrangement of the data derives from creative choices (chiefly case C-604/10 *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* [2012], EU:C:2012:115). This standard is not met when the construction of the database depends on technical rules or constraints that “leave no room for creative freedom” (ibid, para 39). On the basis of Article 3(1) Database, second sentence (“no other criteria shall be applied to determine their eligibility for that protection”), the Court excluded that a mere investment of significant skill and labour could meet the originality requirement, thus preempting the future application of the standard used by UK courts and, albeit in more circumscribed instances, by other national systems (ibid, para 41; Hugenholtz 2012, 54).

The most innovative regulatory solution adopted by the EU, however, consists in the *sui generis* right conferred on the content of the database on ground of the qualitatively and quantitatively substantial investment that the database author had to undertake in order to obtain, verify or present the material (Article 7 Database). The CJEU has already clarified that the right does not protect investment in generating the data or other contents, in light of the economic rationale of the Directive, which is “to promote the establishment of storage and processing systems for existing information, and not the creation of materials capable of being collected subsequently in a database”.<sup>1</sup> Still, the separation line between “creating” and “obtaining” data is thin, and it is prone to create several uncertainties in the field of machine-generated data (Davison-Hugenholtz 2010, 114).

---

<sup>1</sup> Case C-203/02, *British Horseracing Board v William Hill Organization* [2004] ECR I-10497. See similarly Case C-46/02 *Fixtures Marketing Ltd v Oy Veikkaus Ab* [2004] ECR I-10396; Case C-388/02 *Fixtures Marketing Ltd v Svenska Spel AB* [2004] ECR I-10497; Case C-444/02, *Fixtures Marketing Ltd v Organismos prognostikon agonon podofairou AE (OPAP)* [2004] ECR I-10549.

Article 7 Database articulates the entitlement as the right to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the content of that database. “Extraction” entails the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium, by any means or any form. “Re-utilization” means the making available to the public of all or a substantial part of the content of a database by the distribution of copies, by renting, by on-line or other forms of transmission. The same provision specifies that also the repeated and systematic extraction and/or re-utilization of insubstantial parts should be prohibited if the result conflicts with a normal exploitation of that database or unreasonably prejudice the legitimate interests of its maker. Not unexpectedly in light of their relevance and conciseness, most of these concepts have been subject to scrutiny and clarification by the CJEU (for an overview see Derclaye 2012, 343-345).

Despite recitals 45 and 46 Database are clear in stating that the *sui generis* right “does not in any way constitute an extension of copyright protection to mere facts or data”, nor should it “give rise to the creation of a new right in the works, data or materials themselves”, the new entitlement has been accused of increasing the risk of information monopolies due to its close resemblance to a property right on aggregated data (Reichman, 455). This conclusion finds confirmation in the term of protection envisioned for the right. In fact, while the copyright on an original database lasts for 70 years *post mortem auctoris*, not renewable, the *sui generis* right has a duration of 15 years from the 1<sup>st</sup> of January of the year of completion – a term which restarts every time the producer makes a substantial change to the contents of the database, “which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively” (Article 10). Similar concerns have been raised by the Evaluation Report on the Database Directive, which has underlined the flaws and drawbacks of the right, noting how the positive impact the EU legislator expected from Article 7 on the competitiveness of the European database industry has never really materialized (SWD Evaluation 2018, 14). Yet, no intervention on the *sui generis* right seems to be envisioned in the current plans of revision of the Directive (Communication “A European strategy for data” 2020, 13).

A completely different approach has been adopted in the US, where the Copyright Act does not contain a separate provision for databases, which are protected as compilation work under §103, according to which “the copyright (...) extends only to the material contributed by the author of such work, as distinguished from the preexisting material employed in the work, and does not imply any exclusive right in the preexisting material”. General copyright rules apply, such as the idea-expression dichotomy, which excludes from protection abstract ideas, raw data, facts, methods and the like, and the originality requirement, which requires the work to hit a certain threshold of creativity and individuality to be copyrightable (Davison 2003, 160 et seq.). While the bar was originally set very low, for the “sweat of the brow” doctrine demanded only the investment of time and effort to grant protection, the Supreme Court revised this approach in 1991 with the landmark *Feist Publications, Inc. v Rural Telephone Service Co.* (499 U.S. 340), which ruled that a minimum level of creativity in the selection and structuring of the material was requested for databases to be copyrightable. After *Feist*, several decisions have developed the doctrine of database protection ruling, for instance, that the creativity requested had to be referred to the selection of facts and the structure of the work, and not to the means used to discover facts and data (*Bellsouth v Donnelley*, 999 F.2d 1436 (1993)). Case after case, the originality requirement was lowered to include also non-original databases within the scope of copyright. Already *Feist* clarified that “the requisite level of creativity is extremely low; even a slight

amount will suffice. The vast majority of works make the grade quite easily, as they possess some creative spark, "no matter how crude, humble or obvious" it might be". In *Key Publications Inc v Chinatown Today Publishing Enterprises Inc.* (945 F.2d 509, (2<sup>nd</sup> Cir. 1991)), the court argued that database protection may be "thin", but "we do not believe it is anorexic", and the very same Circuit found that the originality standard was satisfied when the compiler's original contribution was subjective and evaluative, distinguishing between soft facts, which are mixed with the author's evaluative judgment and taste and may thus be protected, and hard facts, which are excluded from copyrightability (*CCC Information Services v. Maclean Hunter Market Reports Inc.*, 44 F.3d 61 (2d Cir. 1994)).

Albeit strong and broad in its scope, the protection offered to databases does not go as far as the EU *sui generis* right does. As made clear in *Assessment Technologies v Wiredata* (350 F.3rd 640 (2003)), in fact, database copyright cannot extend to materials in public domain, for the scope of exclusive rights had to be referred only to the selection, organization, verification, classification and presentation of the database content, and not to the content itself. Several bills were tabled to introduce a *sui generis* protection also in the US, but all of them were rejected or got lost in the legislative pipeline (Davison 2007). Several studies have noted, however, how the US legal system offer other protection tools to the database industry, ranging from remedies against misappropriation and unfair competition to contractual instruments, trade secret and trademark enforcement, which together make it possible to create layers of exclusivity the pervasiveness of which is not any less than the intensity of the protection granted by the EU Database Directive (Leaffer, 855 et seq.; Copyright Office 1997, 27).

### **3. Exceptions and limitations to database protection**

As a consequence of the bifurcated structure of protection, in the EU the types of exceptions are different depending on whether the right involved is the copyright on the database structure, or the *sui generis* right on its content (Derclaye 2008, 119 et seq).

Article 6 Database lists the exceptions to copyright, which remains distinct from the general exceptions of Article 5 of the InfoSoc Directive (2001/29/EC, Article 5). The provision introduces a general exception for lawful users, who can perform without authorization of the right-holder any of the restricted acts listed in Article 5 Database, when necessary "for the purposes of access to the contents of the databases and normal use of the contents". Then, it gives Member States the possibility to provide additional limitations for (i) reproduction for private purposes of a non-electronic database; (ii) illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved; (iii) purposes of public security or administrative or judicial procedure; (iv) where other exceptions to copyright provided by national law are involved, without prejudice to the above listed exceptions. The provision is deemed mandatory and any derogation by contract is therefore to be considered null and void (Article 15 Database). Article 9 Database lists, instead, the exceptions to the *sui generis* right, which include (i) the extraction for private purposes of the contents of a non-electronic database; (ii) the extraction for purposes of illustration for teaching or scientific research, to the extent justified by the non-commercial purposes to be achieved and if the source is indicated; and (iii) the extraction or re-utilization for purposes of public security or administrative or judicial procedures.

Recently, the Directive on Copyright in the Digital Single Market (CDSMD) has introduced an additional mandatory exception to Articles 5 and 7 Database to allow text and data mining activities. For more details, see entry *Copyright*.

Countries using only a copyright-based system of protection apply their general copyright exceptions also to databases. In the United States, for example, the fair use doctrine has been applied in the field, with additional specification of its criteria to adapt them to the particular subject-matter involved (among the most recent examples, see *The Author's Guild v. Hathitrust*, 755 F.3d 87 (2d Cir. 2014); *Fox News v. TVEYES, Inc.*, 43 F. Supp. 3d 379 (S.D.N.Y. 2014); for a comment see Samuelson 2014).

## **BIBLIOGRAPHY**

Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as revised at Paris on July 24, 1971 and amended in 1979, S. Treaty Doc. No. 99-27 (1986)

Beutler, S (1996) The protection of multimedia products through the European Community's Directive on the Legal Protection of Databases, *Entertainment Law Review* 325.

Davison, M (2003), *The legal protection of databases*, Cambridge, Cambridge University Press.

Davison, M (2007), Database Protection: Lessons from Europe, *Congress and the WIPO*, 57 *Case Western Reserve Law Review* 829.

Davison, MJ, Hugenholtz, PB (2005), Football fixtures, horseraces and spin-offs: the ECJ domesticates the database right, 27(3) *European Intellectual Property Review* 113.

Derclaye, E (2008), *The Legal Protection of Databases. A comparative analysis*, Cheltenham-Northampton, Edward Elgar.

Derclaye, E (2012), Database rights: success or failure? The chequered yet exciting journey of database protection in Europe, in Geiger, C (ed), *Constructing European Intellectual Property. Achievements and New Perspectives*, Cheltenham-Northampton, Edward Elgar, 340-354.

Dreier, Th (1992), Die Harmonisierung des Rechtsschutzes von Datenbanken in der EG, *GRUR Int* 739.

IDC-Open Evidence (2017), *European Data Market Study. Final Report prepared for the European Commission (SMART 2013/0063)*.

European Commission (2020), *Communication "A European Strategy for Data"*, 19 February 2020, COM(2020) 66 final.

European Commission (2018), *Staff Working Document "Evaluation of Directive 96/6/EC on the legal protection of databases"*, SWD(2018) 147 final.

European Commission (1992), Explanatory memorandum to the Proposal for a Council Directive on the legal protection of databases, COM(92) 24 final.

European Commission, Follow-up to the Green Paper on Copyright in the Information Society, 17 January 1991, COM(90) 584 final.

Ficsor, M (2002), *The Law of Copyright and the Internet: The 1996 WIPO Treaties, their Interpretation and Implementation*, Oxford, Oxford University Press.

Gervais, D (2017), *The TRIPS Agreement: Drafting History and Analysis*, 4<sup>th</sup> ed., London, Sweet and Maxwell.

Goldstein, P, Hugenholtz, B (2010), *International Copyright. Principles, Law and Practice*, Oxford, Oxford University Press.

Hugenholtz, PB (2000), *Implementing the Database Directive*, in Kabel, JC, Mom, GJHM (eds), *Intellectual Property and Information Law. Essays in Honour of Herman Cohen Jehoram*, The Hague, Kluwer, 183-202.

Leaffer, M (2016) *Database Protection in the United States is Alive and Well: Comments on Davison*, 57(4) *Case Western Reserve Law Review* 855.

Masouye, C (1978), *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris act, 1971)*, WIPO, Geneva.

Pattison, M (1992), *The European Commission's Proposal on the Protection of Computer Databases*, 4 *EIPR* 116.

Reichman, JH (2002) *Database Protection in a Global Economy*, 2-3 *Revue Internationale de Droit Economique* 455.

Ricketson, S, Ginsburg, JC (2006), *International Copyright and Neighboring Rights: The Berne Convention and Beyond*, Oxford, Oxford University Press

Rosati, E (2013) *Originality in EU Copyright. Full Harmonization through Case Law*, Cheltenham-Northampton, Edward Elgar.

*TRIPS: Agreement on Trade-Related Aspects of Intellectual Property Rights*, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994)

Samuelson P (2014), *Mass Digitisation and Fair Use*, 57(3) *Communications to the ACM* 50.

Seuba, X, Geiger, C, Pénin, J (2018), *Global Perspectives and Challenges for the Intellectual Property System*, CEIPI-ICTSD, Issue Number 5, Geneva-Strasbourg



US Copyright Office (1997), Report on Legal Protection for Databases: A Report of the Register of Copyrights, August 1997, Washington D.C., Government Office Printing Staff.

von Lewinski, Silke (2008), International Copyright Law and Policy, Oxford, Oxford University Press.

Walker, MM, W, von Lewinski, S (2010), European Copyright Law. A Commentary, Oxford, Oxford University Press.

Wiebe, A (2004) Database Protection in Europe in the Aftermath of *William Hill* and *Fixtures*, MR Int 38.

WIPO Copyright Treaty, 36 ILM 65 (1997); WIPO Performances and Phonograms Treaty, 36 ILM 76 (1997).

WIPO-UNESCO Second Committee of Governmental Experts (1982), Copyright Problems Arising from the Use of Computers for Access to or the Creation of Works, 7-11 June 1982, reprinted in 18 Copyright 239.