

# Open licensing and databases

*Simone Aliprandi,<sup>a</sup>*

*(a) Lead of the Copyleft-Italia.it Project,  
member of Array (Arraylaw.eu) and*

*Ph.D. in Information Society at Bicocca University of Milan*

DOI: [10.5033/ifosslr.v4i1.62](https://doi.org/10.5033/ifosslr.v4i1.62)

## **Abstract**

Data and databases are a complex, nuanced area within intellectual property law.

In the European Union databases have a special legal treatment that provides two levels of protection. A database is protected by copyright in the classical sense when it can be considered an intellectual work with a creative nature. Where databases represent mere collections of data without sufficient creativity to trigger copyright, EU jurisdictions protect the database under *sui generis* rights when substantial investment has been made in obtaining, verifying, or presenting the database contents according to Directive 96/9/EC.

This system creates a substantial discrepancy between the situation of European countries and the rest of the world, and also affects those databases that have been released under open licenses.

Not all of the currently available open licenses take account of the legal and practical implications of this discrepancy, and we should examine the consequences and options.

The paper aims to provide a high-level analysis on the protection of databases under European law and identify the main legal problems arising from it in an open data scenario. Then it will focus on the solutions tried so far to implement a proper open licensing framework for the database (with an introduction to the licenses offered by Creative Commons and the Open Data Commons project). Finally, some of the most prominent use cases of open licensing for data will be analysed (such as those of geo-data and linked-data), with some observations on the modus operandi of the various promoters of projects.

## **Keywords**

Open data, open licensing, open content, public domain, Creative Commons, copyright, database right.

## 1. Introduction: data and database

As is well-known, digital technologies allow the management, storing and processing of huge amounts of information. Work that recently required the contribution of many people can now be done using a simple automated software; information which had to be stored in entire rooms a few years ago can now be stored on a very small USB pen; tasks that once required entire working days to complete can now be easily sorted out in few minutes. Time, space and effort have been reduced, to the benefit of a constantly increasing supply of data and increasingly numerous ways of managing it.

But what exactly is “data”?

*“The word data is the Latin plural of datum, neuter past participle of dare, “to give”, hence “something given”. [...] Also, data is a representation of a fact, figure, and idea. Such usage is the origin of data as a concept in computer science: data are numbers, words, images, etc., accepted as they stand”.<sup>1</sup>*

It may seem obvious but, in order to avoid dangerous misunderstandings, I think it is important to clarify the meaning of “data”; there is confusion about the real meaning of this term. Indeed, there is a trend of generally talking about “data” when referring to all the material stored on a computer or digital media, regardless of whether it is films, music files, documents, images etc.

From the point of view of legal language (which must be taken into consideration when making an observation of this kind) “data” has a smaller semantic range and only refers to “facts” which are not organized and processed by human intelligence. These, as single pieces of information deducible from the nature of things, are not subject to copyright protection and patent rights, and are therefore not important from the point of view of the right of intellectual property.

Intellectual property does not deal as much with data as it deals with databases, and it is very important to always consider this distinction.

According to the Encyclopedia Britannica, in the computer science field a database is «any collection of data, or information, that is specially organized for rapid search and retrieval by a computer». <sup>2</sup> Therefore, “data” is only subject to regulations and legal protection when it is presented as organized systems. <sup>3</sup>

As will be seen below, with the advent in the 1990s of an *ad hoc* European regulation for database protection, the concept of the database has been further clarified and explored by legal science. <sup>4</sup>

Obviously, it is no coincidence that a need for questioning the appropriateness of a particular legal process for databases has only arisen in recent decades: this is closely linked to the new

1 <http://en.wikipedia.org/wiki/Data>.

2 <http://www.britannica.com/EBchecked/topic/152195/database>.

3 Of course, we have always to consider the existence of other legal protection, such as the rules related to trade secrets and unfair competition.

4 See, in this regard, the definition provided by art. 1.2. of directive 96/9/EC: «'database' shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.»

possibilities for collecting, organizing and using huge amounts of data stemming from digital technologies and business opportunities based on this kind of activity.

## 2. The particular legal treatment for databases in Europe

### 2.1. Before the Database Directive

In a way, databases can be compared to collective works, a category recognized in the copyright field long before the reforms of the 1990s. Indeed, the Berne Convention and, in general, all national regulations inspired by it, also include, among the types of works protected by the law, those created through the collection of other works independent from the collective work.

Indeed, this is the text of art. 2 (5) of the Berne Convention: «Collections of literary or artistic works such as encyclopedias and anthologies which, by reason of the selection and arrangement of their contents, constitute intellectual creations shall be protected as such, without prejudice to the copyright in each of the works forming part of such collections».<sup>5</sup>

The person who selects, collects and organizes data according to particular creative criteria holds, therefore, a stand-alone copyright with respect to the individual collected works.

With the advent of new methods of storage and technological management of information, databases have become a fundamental part of cultural and technical production. Therefore, the world of law has begun questioning whether specific forms of protection for this new category of creations are necessary or if, on the contrary, it is enough to (extensively) apply pre-existing copyright categories and principles.

### 2.2. The inadequacy of the classic copyright protection

From a first reading of the overall copyright principles, it can be easily grasped that the definition of collective works (in the sense of collection of works) refers to phenomena not always comparable to a database. Not all databases have a requirement of choice and organization of material according to creative criteria; “particularly not those that, offering to provide all available information about a certain topic, do not apply any form of selection and present the information itself in a manner that is either uninteresting or dictated by information requirements”.<sup>6</sup>

Furthermore, there is another “Achilles Heel” with regard to the copyright of atypical works such as databases: namely the principle that the copyright only covers the expressive form of a work, that is, the way the author expresses their idea and not the idea itself. Therefore, and particularly in this case, on the basis of the sole copyright, another person may use the contents of the database,

---

5 [http://www.wipo.int/treaties/en/ip/berne/trtdocs\\_wo001.html](http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html). Also see, in this regard, Article 5 of WIPO Copyright Treaty, 1996: «Compilations of Data (Databases) – Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. This protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation.»

6 AUTERI, P., *Diritto d'autore*, part VI of *Diritto industriale. Proprietà intellettuale e concorrenza*, Giappichelli, 2005 (pp. 505-508).

modifying the organization and arrangement criteria, effectively creating a work which is different from a legal point of view, but substantially repetitive and “parasitic”.

With the sole application of the copyright, a large portion of databases would be left without any legal protection; all that would remain would be protection ensuing from the principle of unfair competition or the possible application of technological protection systems. This has been considered insufficient by the European legislator, who, after a lively debate on the appropriateness of the choice, decided to take action with a special directive.

This choice has been supported by the idea according to which certain types of databases that would be excluded, because of their nature, from the scope of the copyright, require a large investment and that, therefore, this investment in itself is worthy of protection and should be encouraged accordingly.<sup>7</sup>

### 2.3. A double level of protection: the EU Directive and the sui generis right

Therefore, in 1996, the European legislator decided to outline a special model of protection, according to which databases are potentially eligible for a double level of protection. According to Directive n. 96/9/EC, on the one hand, databases have been formally included among the categories of creative works protected by copyright in the community legislation; on the other hand, special rights have been created for the author of the database. As Paolo Auteri points out:

*«the object of the first protection [copyright] is the “expressive form”, i.e. the way information material is selected and organized, while the object of the second protection is the information content or rather the information as a whole, in the measure in which research, verification and presentation require a significant investment».*<sup>8</sup>

The text of the Directive is made up of sixteen articles divided into four Chapters. Chapter II deals with the protection of databases as the author's own intellectual creation<sup>9</sup> and therefore to be protected by copyright. Up to this point, the Directive does nothing more than clarify and formally ratify what was already easily inferable from the principles of copyright.

Conversely, the truly innovative (and also the most criticized) part of the Directive is Chapter III, where new rights are established for protecting databases of a non-creative type, which are therefore not considered legitimate intellectual works. Such rights (generally referred to with the Latin phrase “*sui generis* right”, in order to highlight their peculiarity with respect to copyright and related protection rights) are exclusive rights arising from a party referred to by the regulation

---

7 Read in this regard the Whereas n. 7 and n. 12 of the Directive: 7) Whereas the making of databases requires the investment of considerable human, technical and financial resources while such databases can be copied or accessed at a fraction of the cost needed to design them independently; 12) Whereas such an investment in modern information storage and processing systems will not take place within the Community unless a stable and uniform legal protection regime is introduced for the protection of the rights of makers of databases.

8 AUTERI, P., *Diritto d'autore*, part VI of *Diritto industriale. Proprietà intellettuale e concorrenza*, Giappichelli, 2005 (pp. 505-508).

9 Art. 3.1: In accordance with this Directive, databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright. No other criteria shall be applied to determine their eligibility for that protection.

as “the maker of the database”; they relate to the money invested in the creation of the database (and not to the creative contribution, as in the case of copyright and related rights protection) and are valid for 15 years from the constitution of the database<sup>10</sup>. The principles of the Directive were then implemented by the member states of the EU and have become an integral part of national regulations, thus standardizing to a certain extent the regulatory framework of the European Union.<sup>11</sup>

Chapter III, which deals with the *sui generis* right, describes two primary activities for which the “maker” is responsible and over which these rights are exercised: the extraction of data from the database (understood as the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form) and the re-utilization of data (understood rather as any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission).<sup>12</sup>

In other words, the maker has the exclusive right, for a period of 15 years, to control these activities on the database (or on a substantial part of it) that they created and made available to the public. This – precisely – occurs in the case of a database without creative features, but which has required a substantial investment in terms of quality and quantity.

#### 2.4. Database categories according to protection levels

As a result of the principles established by the Directive and therefore the different cases of overlap between the two levels of protection, it is possible to outline the following categories of databases protected by the European regulation:

- *Type 1* - Databases with creative features containing creative works
  - protected by copyright on two independent levels
    - the author of the database holds the copyright with regard to its structure and the specific organization of its contents; the authors of the individual contents hold the copyright on the independent contents in a totally independent manner.
- *Type 2* - Databases with creative features containing simple data
  - protected on two different levels (copyright and *sui generis* right)
    - the author of the database holds the copyright with regard to its structure and the specific organization of its contents; the author themselves also fills the role of maker and holds the *sui generis* right as far as the extraction and re-utilization of substantial parts of the data are concerned.

---

10 Specifically, art. 10.1 of the Directive reads: «The right provided for in Article 7 shall run from the date of completion of the making of the database. It shall expire fifteen years from the first of January of the year following the date of completion.»

11 In 2005 the European Commission has published an evaluation of the protection EU law gives to databases. This interesting and insightful report is available at [http://ec.europa.eu/internal\\_market/copyright/docs/databases/evaluation\\_report\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf).

12 See Art. 7 – Object of protection.

- *Type 3* - Databases without creative features containing simple data, but nevertheless requiring a significant investment
  - protected only by the *sui generis* right
    - the maker of the database holds the *sui generis* right as far as the extraction and re-utilization of substantial parts of the data are concerned.<sup>13</sup>

This pattern highlights how important it is that the two levels of protection are always clearly defined, especially when dealing with the licensing of a database.

We should always have very clear ideas about what rights and what objects we intend to license; at the same time, we should try to clearly communicate our intentions to the licensees, specifying whether we are referring to the database itself, its contents or both.

The determining factor for the subdivision in these three types – as often occurs in copyright – is the presence of creative features. A study on this concept would require a number of pages and it is not possible to look further into the matter here; therefore, you are advised to refer to more specialist sources and study case law on database protection.<sup>14</sup>

### 3. The open licensing paradigm applied to databases

#### 3.1. Licenses that do not license

Once the complexity of this protection system has been clarified, it is possible to deal with the problems that arise when the holder of the rights on a database decides to regulate its use through the application of a free distribution license or copyleft.

As already pointed out, all the most common licenses that one would consider to also license databases (such as GPL, GFDL, Creative Commons) are modelled upon a “pure” copyright system. This does not always mean they conveniently deal with the *sui generis* right, which differs in some aspects from copyright (in the strictest sense of the word). Therefore, their use in the field of databases in the European area may not cover the part relative to the *sui generis* right.

Let us try to understand this better. The function of these licenses is to authorize, permit or, more precisely, “license” free use of the work to which the license refers, and in order to do so, the text of the licenses explicitly refers to the single rights involved in the cession. However, not all these licenses expressly take into consideration the *sui generis* right.

There is a reason for this: most of these licenses, despite having been “exported” to Europe, were conceived within the US legal system, where the double protection level for databases does not

<sup>13</sup> We can also find more complex cases of databases, with hybrid features or made by the ensemble of other (already existing) databases.

<sup>14</sup> «This feature may be sought alternately or cumulatively in the choice or arrangement of materials.» UBERTAZZI, L.C. (editor), *Diritto d'autore, estratto da Commentario breve alle leggi su Proprietà Intellettuale e Concorrenza*, 4° ed., CEDAM, 2009 (p. 185)

exist.

Essentially, whenever we have to deal with a database licensed under one of these licenses, we cannot be assured to be able to use it freely as, except in the case a specific integration to the license text is added, the rights holder (i.e., the maker) would withhold the full control over the *sui generis* right.

It is, therefore, necessary to think of the best way to deal with these particular types of rights and there are substantially two ways: either the waiver of these rights, or their specific licensing.

An important clarification: the considerations below refer only to the licensing of databases not considered intellectual works and therefore only protected by the *sui generis* right (i.e. the Type 3 described in paragraph 2.4).

### 3.2. The waiving option

The first of the two ways that can be implemented involves the maker waiving their rights on the database, before the first 15 years foreseen by the Directive have elapsed and the database permanently enters the condition of *public domain*.

In order to reach this situation, it is necessary for the holder of the rights to issue a public statement in which they waive their rights in an unlimited and unconditional manner.

This solution has been successfully applied with reference to copyright: consider the Public Domain dedication proposed in the past by the Creative Commons project and the latest tool called CC zero, which allows the holder of the copyright to release their work in a sort of artificial public domain.<sup>15</sup> A similar outcome may be reached through the Public Domain Dedication and License (PDDL) proposed by the Open Data Commons Project.<sup>16</sup>

On the one hand, this solution certainly provides the most freedom of use of the database and, on the other, creates fewer problems from the point of view of the distribution and use of the database on an international scale. Indeed, in the event of a maker of a European database<sup>17</sup> deciding to waive the *sui generis* right, they would allow their product to circulate freely without any doubts about the systems of protection to be applied. Accordingly, a non-European user need not wonder whether the database, coming from the European area, is protected by a protection system that differs from the one used in their own country.

The approach of waiving the *sui generis* right was promoted by Creative Commons, not only with a call to use the CC0 tool as much as possible for the release of databases<sup>18</sup>, but also by adding a

---

15 Further details on tools proposed by Creative Commons for the public domain can be found on the website <http://creativecommons.org/publicdomain/>.

16 In this regard see the website [www.opendatacommons.org/licenses/pddl/](http://www.opendatacommons.org/licenses/pddl/).

17 Database rights are only granted to European makers. This is a considerable difference with the copyright, which is granted in Europe – as in any country who adopted the Berne Convention – regardless of the country of first publication.

18 «We do recommend CC0 for scientific data — and we're thrilled to see CC0 used in other domains, for any content and data, wherever the rights holder wants to make clear such is in the public domain worldwide, to the extent that is possible (note that CC0 includes a permissive fallback license, covering jurisdictions where relinquishment is not

specific waiver in European version 3.0 licenses.<sup>19</sup>

As a matter of fact, the Creative Commons licenses, in their original conception, do not expressly take into consideration the so-called “database rights” as they represent licenses originating from the United States.<sup>20</sup> Those licenses had become one of the main points of reference for the free distribution of intellectual works in the European countries; it would have been therefore problematic for their full adoption had they been left outside the field of application of databases.

Therefore, the need was perceived to adapt the national versions so that they could also license the *sui generis* right. This porting process required long deliberation and a constant comparison between the various national working groups of the Creative Commons Project, and it was only completed at the beginning of 2011, with the release in these countries of the version 3.0 license, in which “database rights”<sup>21</sup> have been specifically mentioned and waived.<sup>22</sup>

### 3.3. The specific licensing option

The waiver solution is not always applicable and therefore due licensing of the *sui generis* right is required. It refers, for example, to those cases in which the holder of the rights intends to release the database with specific conditions, such as, for example, the attribution of authorship or the so-called “share-alike”. In these cases, it would only have an effect on the *sui generis* right.

In 2008, an independent project also got underway, aimed at the creation of a license specifically designed for databases.<sup>23</sup> This British project began life at Edinburgh University, through the initiative of a Texan lawyer who had moved to Scotland to continue his work as a researcher and teacher: Professor Jordan Hatcher<sup>24</sup>. The most important result of this project, which by no accident was called “Open Data Commons”, was the release of the license “Open Database License” (OdbL)<sup>25</sup>.

The ODbL is a rather complex but well put-together license; and it can effectively apply the copyleft model with reference to databases. It includes, in fact, a set of clauses that reproduce the model of the Attribution – Share Alike licenses proposed by Creative Commons.

It licenses only the right relating to databases; therefore, if the database contains creative works, in

---

thought possible).» <https://creativecommons.org/weblog/entry/26283>.

19 «We adopted a policy that version 3.0 EU jurisdiction ports must waive license requirements and prohibitions (attribution, share-alike, etc) for uses triggering database rights — so that if the use of a database published under a CC license implicated only database rights, but not copyright, the CC license requirements and prohibitions would not apply to that use.» <https://creativecommons.org/weblog/entry/26283>.

20 This approach is however likely to change with the upcoming release of the version 4 of the Creative Commons Licenses, still under development at the time of writing.

21 On the relationship between Creative Commons licenses and database rights read the study <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>; and also the page <http://sciencecommons.org/resources/faq/database-protocol/>.

22 See for example the Italian porting of CC 3.0.

23 The full text of the law is available on the website <http://www.opendatacommons.org/licenses/odbl/>

24 Hatcher's personal blog has quite an emblematic name: <http://www.opencontentlawyer.com>.

25 Some of the activists involved in this project had previously dealt with another license of the same type, in truth rather superficial and almost immediately abandoned: the Talis Community License, currently available on the website <http://w.talis.com/tdn/tcl>



order to guarantee free use of the whole work, it is advisable to apply another license relative to the works contained in the database itself. Indeed, the preamble of the license specifies as follows: «Because databases can have a wide variety of types of contents, this document only governs the rights over the database, and not the contents of the database individually. You should use the Open Data Commons together with another license for the contents, if the contents have a single set of rights that governs all of them». This implies the need for a certain degree of shrewdness in choosing the license for the content: so as not to create further complications for licensees and indeed also for interpreters (lawyers, judges...), it is necessary to choose a license that reproduces the same effects for the contents as well.

Between the choice of waiving the *sui generis* right and the choice of licensing with the share-alike clause there is obviously an intermediate option, that is a licensing that only requires the attribution of authorship of the original database. In essence, it is the same effect produced by a Creative Commons Attribution brought in the scope of the mere *sui generis* right rather than the copyright.

For this purpose, the Open Data Commons project has proposed a further license called a “ODC Attribution License”<sup>26</sup>.

From its own point of view, while still following this pattern, the government of the United Kingdom drew up a specific license called “License to Reproduce Public Sector Information”, the aim of which is to release, in open mode, information (by which is meant both content and data) produced by British institutions where so-called Crown Copyright is applied<sup>27</sup>. There is a recent evolution of this project: a new “attribution-only” license delivered by The National Archives and called “Open Government Licence for public sector information”.<sup>28</sup> The choice of creating this license by the British government is part of the pursuit of the objectives set by the European Union with the Directive 2003/98/EC on the re-use of public sector information<sup>29</sup> and the subsequent Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)<sup>30</sup>.

26 The foreword in the license reads: «The Open Data Commons Attribution License is a license agreement intended to allow users to freely share, modify, and use this Database subject only to the attribution requirements set out in Section 4.» The complete text of the license is available at [www.opendatacommons.org/licenses/by/](http://www.opendatacommons.org/licenses/by/).

27 Indeed, the article 1.3 of the license reads: «Waived material can be re-used free of charge without requiring a formal license provided that it is: i) acknowledged; ii) not used in a misleading way; iii) reproduced accurately and kept up to date». The full text of the document is available at [www.opsi.gov.uk/click-use/system/licenceterms/CCWPS03-00.pdf](http://www.opsi.gov.uk/click-use/system/licenceterms/CCWPS03-00.pdf). Although the database rights are not expressly mentioned, it is clear from the context and scope of the license that it deals with *sui generis* rights in the first place.

28 <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

29 «Directive 2003/98/EC on the re-use of public sector information, otherwise known as the PSI Directive is an EU directive that encourages EU member states to make as much public sector information available for re-use as possible. Previously this area was left to member states to regulate. This directive now provides a common legislative framework for this area. The Directive is an attempt to remove barriers that hinder the re-use of public sector information throughout the Union.» [http://en.wikipedia.org/wiki/Directive\\_on\\_the\\_re-use\\_of\\_public\\_sector\\_information](http://en.wikipedia.org/wiki/Directive_on_the_re-use_of_public_sector_information)

30 In the proposal act of the Directive (par. 2) we can read: «The proposed Directive creates a legal framework for the establishment and operation of an Infrastructure for Spatial Information in Europe, for the purpose of formulating, implementing, monitoring and evaluating Community policies at all levels and providing public information. A key objective of INSPIRE is to make more and better spatial data available for Community policy-making and implementation of Community policies in the Member States at all levels. INSPIRE focuses on environmental policy but is open for use by and future extension to other sectors such as agriculture, transport and energy.» [http://ec.europa.eu/information\\_society/policy/psi/docs/pdfs/inspire/en.pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/inspire/en.pdf).

The other European countries are trying to move in the same direction set by the two directives, with different initiatives for the issuing of public data<sup>31</sup>. In France a license called Licence Information Publique, whose effects are similar to the Creative Commons Attribution license, has been released<sup>32</sup>. In Italy, one of the most interesting pilot projects is the one set up by the Piedmont Region, which has chosen to primarily use the CC0<sup>33</sup> tool; recently also a new license has been created and called Italian Open data license (with version 1.0 following the CC by-sa paradigm and version 2.0 following the CC by paradigm)<sup>34</sup>.

As is clear by reading this paragraph, which has a purely introductory purpose and does not get to the heart of the matter concerning the emerging legal issues, the choice of a specific licensing of a database protected by the *sui generis* right implies some considerable legal complications.

A clear overview of all the most important licenses cited above is available as a diagram where they are classified according to their legal effects (attribution and share-alike, attribution only, public domain).<sup>35</sup>

## 4. Some interesting cases

### 4.1. Openness in geodata: the Open Street Map project

One of the most interesting cases to have dealt with this kind of problem is the extremely topical one of geographical data and its use in an open pattern. In the wake of cultural movements inspired by the free sharing of contents (open source, open content, open access), a growing share of activists/volunteers have become committed to the creation of a geographic information system (the so-called GIS) that is freely accessible and usable, without being subject to intellectual property restrictions.

On the other hand, when talking about the relationship between databases and open licensing, this topic cannot be overlooked, as it was precisely because of the cultural ferment stemming from communities developing free geographical data that the importance of also delving into certain aspects from a legal viewpoint was perceived.

The category of geographical data is difficult to qualify from a legal point of view, as it concerns various kinds of creativity and representations of reality.

We may have to deal with “simple” data such as, for example, coordinates of longitude and latitude, height, distance from points of interest etc.; and in this case the single data item certainly

---

31 For a complete overview of the main projects inspired by the “open data” model in Europe see the interesting study “Open Data, Open Society” carried out by Marco Fioretti for Scuola Sant’Anna di Pisa (available on the website [www.dime-eu.org/node/907](http://www.dime-eu.org/node/907)).

32 More information about the license and its entire text are available at [http://www.rip.justice.fr/information\\_publique\\_librement\\_reutilisable](http://www.rip.justice.fr/information_publique_librement_reutilisable).

33 The official website of this project is <http://dati.piemonte.it/>.

34 Version 1.0 is available at <http://www.formez.it/iodl/> and Version 2.0 is available at <http://www.dati.gov.it/iodl/2.0/>.

35 The diagram is available at [http://www.ifosslr.org/public/opendata\\_graph.pdf](http://www.ifosslr.org/public/opendata_graph.pdf) or at [http://www.aliprandi.org/doc/opendata\\_graph.pdf](http://www.aliprandi.org/doc/opendata_graph.pdf).

cannot be protected by copyright, as it would be nothing more than a natural “fact”, a revelation of reality, without any mediation by the human mind. As already explained, this kind of data can only be protected as an “organized system of data”, through the *sui generis* right.

If, on the other hand, we have to deal with something more elaborate and, above all, that has required a certain creative approach, the situation becomes more complicated.

In this case, in order to assess what level of protection to apply to the contents, it is necessary to verify, each time, the type of creative work (among those envisaged by the principles of copyright) that the reprocessed and conceivably represented data item may be included in. It is not always an easy analysis to undertake, as contents sometimes appear in the form of aerial or satellite photographs (protected by a relevant right); other times (and this is currently the most frequent case), they are not real photographs but (two-dimensional or three-dimensional) vector graphic reconstructions of a geographical reality, and therefore more likely to be assimilated to architectural and engineering works (drawings, projects, etc.) and thus also protected by a relevant right.

There are also those who have pointed out that a map containing georeferenced information (e.g. height, average temperatures, frequency of rainfall, texture of the soil, etc.) also implicitly represents a database that is subject to the *sui generis* right as well. This keen observation, however, somewhat complicates the legal qualification of the cartography.

At any rate, in addition to the licenses of wider application analyzed above, several licenses specifically conceived for geographical data have been drawn up in Europe in recent years, the most important of which are listed below.

The first is the Public Geodata License: of French origin, available in French since 2003 and currently also available as an English version dating back to June 2004. Its foreword reads: «This license applies to geographical data, attributes, and associated metadata. It applies to any derivative work, too. Its purpose is to facilitate production, exchange, and distribution of geographical data, in respect of author rights and users rights to benefit of the same liberties.» However, this license, specifically inspired by the principles of GNU GPL (including the copyleft effect), has remained at version 0.1, dating back to 2004, and therefore probably stopped at a purely experimental level<sup>36</sup>.

Between 2004 and 2005, some British activists in the sector proposed to apply a license modeled on the Creative Commons Attribution-ShareAlike and called “Open Geodata License”. The project must have been abandoned at a very early stage (probably because of previously acknowledged legal problems), so much so that only a few traces of the license exist on the web<sup>37</sup>.

There is no doubting the fact that the most relevant current project concerning open geographical data is OpenStreetMap, both in terms of the number of users and active participants and the level of articulation and efficiency of information (data, maps and integrated services) produced by the project.

---

36 See the website [http://en.giswiki.org/wiki/Public\\_Geodata\\_License](http://en.giswiki.org/wiki/Public_Geodata_License).

37 One of the few websites where it is possible to read the document is <http://socialtapestries.com/outcomes/index.html>.

At the date of writing, the geographical data of the OpenStreetMap project are released to the public with a Creative Commons Attribution-ShareAlike 2.0<sup>38</sup>; however a process of transition has already begun and almost completed to the Open Database License 1.0, which the OpenStreetMap Foundation considers more secure from a legal point of view<sup>39</sup>.

In conclusion, when thinking about the legal nature of geographical data, one of the main doubts held by jurists about the effectiveness of the *sui generis* right arises. Going back to our introductory arguments on the principles of the 1996 Directive, let us remember that the *sui generis* right does not protect data itself, but rather data that is collected and organized in a database; and, above all, it comes into play when the creation of the database requires a significant investment.

Right now, 15 years since the approval of the Directive, one may wonder whether current technology has also affected the idea of “significant investment” as initially intended by the European legislator. Nowadays, in fact, any mobile phone equipped with GPS (that is, most of the devices available on the market) can be used to detect high precision geographical data which, thanks to freeware, can also be easily processed, even by non-professional users and enthusiasts<sup>40</sup>.

Therefore, no particular investments seem to have been made in the collection, verification or presentation of data, only in the maintenance of the servers and the usage and sharing platform managed by the Open Street Map Foundation. Can this be considered a sufficient level of investment for *sui generis* protection to be applicable? Let us deliberately leave the question open.

#### 4.2. Wikipedia as a database? The DBpedia project

DBpedia is an interesting project aimed at extracting and restructuring information from Wikipedia so that it is more easily integrated with the so-called “semantic web” or “web of data”<sup>41</sup>.

The project activists carry out a very meticulous job of extracting data from the various Wikipedia entries and then organizing it according to predefined standards, compatible with typical patterns of the semantic web. In this way, the information from individual entries becomes more easily interpretable by on-line computers (in a word, it becomes “machine-readable”)<sup>42</sup> and can therefore

38 Besides the license mentioned (representing the document whereby OSM geographical data is distributed to the public), it is important to consider the “Contributor Terms” which, on the other hand, represent the terms whereby the active participants in the project agree to waiving the data they have collected. The foreword of this document reads: «This contributor agreement (the “Agreement”) is made between you (“You”) and The OpenStreetMap Foundation (“OSMF”) and clarifies the intellectual property rights in any Contents that You choose to submit to the Project in this user account. Please read the following terms and conditions carefully and click either the “Accept” or “Decline” button at the bottom to continue.» [http://www.osmfoundation.org/wiki/License/Contributor\\_Terms](http://www.osmfoundation.org/wiki/License/Contributor_Terms).

39 «We are considering changing to the Open Database Licence (‘ODbL’). This is very similar in intent to our current license, but the OSM Foundation believes it is more secure legally, and offers more clarity for both contributors and users.» [http://wiki.openstreetmap.org/wiki/Legal\\_FAQ#What.27s\\_this\\_about\\_a\\_licence\\_change.3F](http://wiki.openstreetmap.org/wiki/Legal_FAQ#What.27s_this_about_a_licence_change.3F)

40 This is precisely the spirit upon which most of the Open Street Map project is based, with volunteers privately collecting, processing and sharing data according to the guidelines of the project.

41 «The Semantic Web is a “web of data” that enables machines to understand the semantics, or meaning, of information on the World Wide Web.[1] It extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other, enabling automated agents to access the Web more intelligently and perform tasks on behalf of users» [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web).

42 According to a famous statement by Tim Berners-Lee dating back to 1999, this is the spirit of the semantic web: «I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content,

be easily reused in other works.

At the date of writing, the official website of the project carried the following laconic disclaimer:

*“DBpedia is derived from Wikipedia and is distributed under the same licensing terms as Wikipedia itself. As Wikipedia has moved to dual-licensing, we also dual-license DBpedia starting with release 3.4. Data comprising DBpedia release 3.4 and subsequent releases is licensed under the terms of the Creative Commons Attribution-ShareAlike 3.0 license and the GNU Free Documentation License. Data comprising DBpedia releases up to and including release 3.3 is licensed only under the terms of the GNU Free Documentation License.”<sup>43</sup>*

The main doubt that may arise in the mind of a jurist is whether such a short disclaimer is sufficient to clarify the legal status of the DBpedia database. Indeed, both the Creative Commons Attribution-ShareAlike license and the GNU Free Documentation License are licenses not specifically for databases but for creative contents (and rightly so for a work of this type).

When only information in the form of data is extracted from that work in order to be included in a database, the only feasible protection becomes that of the *sui generis* right.

It is far from certain that such activity of extraction implies a derivative relationship in the most technical sense, unless Wikipedia itself is also considered a database covered by a *sui generis* right. But, if this was the case, we would again fall into an impasse which a Creative Commons License in itself would not be sufficient to resolve.

This doubt may exist all the same, even if we apply the opposite approach. Would a Wikipedia entry created using data drawn from a database protected by the *sui generis* right establish a derivative relationship with it? Take the case of an entry referring to a river or mountain, the georeferenced data for which is drawn entirely from the Open Street Map database. In the case of single entries the problem would not even arise, as the quantity of extracted data would not be sufficient to trigger the prohibition concerning the extraction and reutilization of data required by the Directive<sup>44</sup>. But the situation would be quite different if, for example, it occurred with all the rivers or mountains in a specific area. In this case, would a Wikipedia entry created in this way have a derivative relationship with the Open Street Map database? If the answer is yes, a problem of compatibility between the licenses for the two projects would arise.

As can be seen, in both the case of open georeferenced systems and that of DBpedia, the issue becomes intricate and genuinely complex; and, in the end, does nothing more than highlight the weak points of an extremely unclear right such as the *sui generis* right.

---

links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.»

43 <http://wiki.dbpedia.org/Imprint>

44 As seen, in the text of the Directive, reference is in fact made to “all or a substantial part of the contents of a database”. According to Italian law (Court of Catania 8-1-2001), a non-substantial part of a database is represented by an insignificant percentage of the data contained therein (quantitative criterion), which does not present systematic coordination therein (qualitative criterion), so that it cannot, per se, be defined and used as a database, and the reproduction and distribution of which is totally insufficient for devaluing the database protected by the *sui generis* right.

## About the author

*Simone Aliprandi is an Italian attorney-at-law and independent researcher who is constantly engaged in writing, teaching and consulting in the field of copyright and ICT law. He has an additional degree in Public Administration Science and a Ph.D. in Information Society at the Bicocca University of Milan. He founded and still coordinates the Copyleft-Italia.it project and has published several books devoted to openculture and copyleft. He also collaborates as a legal consultant with Array (<http://www.arraylaw.eu>). More information about his activities and further publications: <http://www.aliprandi.org>.*

### Licence and Attribution

This paper was published in the International Free and Open Source Software Law Review, Volume 4, Issue 1 (March 2012). It originally appeared online at <http://www.ifosslr.org>.

This article should be cited as follows:

Simone Aliprandi (2012) 'Open licensing and databases', *IFOSS L. Rev.*, 4(1), pp 5-18  
DOI: 10.5033/ifosslr.v4i1.62

Copyright © 2012 Simone Aliprandi

This article is licensed under a Creative Commons Attribution - NoDerivs 2.0 England and Wales licence, available at <http://http://creativecommons.org/licenses/by-nd/2.0/uk/>

As a special exception, the author expressly permits faithful translations of the entire document into any language, provided that the resulting translation (which may include an attribution to the translator) is shared alike. This paragraph is part of the paper, and must be included when copying or translating the paper.

