

Study Guidelines

by Jonathan OSHA, Reporter General
Anne Marie VERSCHUUR, First Deputy Reporter General
Ari LAAKKONEN, Second Deputy Reporter General
Guillaume HENRY, Ralph NACK and Lena SHEN
Assistants to the Reporter General

2020– Study Question

Rights in Data

Introduction

- 1) In recent years, the amount of data created, recorded, collected and used all around the world has exploded, and in very diverse fields (automotive, health, building, banking, marketing, etc.).

This phenomenon is due, among others, to the development (i) of technologies to record and process data (e.g. sensors, computers) and (ii) of electronic communication and platforms.

One of the major consequences of the accumulation of a huge corpus of data is the development of artificial intelligence (AI) and machine learning, because “*data is the raw material of AI and the emergence of new uses and applications depends on it.*”¹

Data has become so important that it is has been considered as an infrastructure².

Why AIPPI considers this an important area of study

- 2) An increasing number of modern data applications give rise to legal questions about the protection of data, both in terms of protecting one’s *own rights* and *investments*, and in terms of *avoiding infringement* of rights of others.

For instance, *data mining processes* can be expensive and generate extremely valuable outcomes, but often rely on uses of previously existing data that may be difficult to square with existing exceptions and limitations of IP rights.

¹Cédric Villani, *For A Meaningful Artificial Intelligence*, p. 21,
https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

²OECD, 2015, *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, p. 181: “*The economic properties of data suggest that may be considered as an infrastructure or infrastructural resource*”
https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page180

- 3) At the moment, there is uncertainty about rights in data, because in most jurisdictions the legislation doesn't give clear answers to two major questions: Who owns the data (mere data and database)? Who can access the data?
- 4) The lack of harmonisation creates legal uncertainty around the use and exploitation of data collections, which can deter investment and innovation.

Furthermore, uncertainties around ownership can be a barrier to effective trade and transfer of data.

- 5) Given the increasing role and value of data in all forms of innovation across industries and countries, it may be time for this to change.
- 6) However, this does not necessarily mean increasing the overall scope of protection. Although protection may encourage investment, there are also valid concerns about stifling competitive innovation by creating new monopolies on information.

Relevant treaty provisions

- 7) The TRIPs Agreement contains provisions on the protection of (i) compilations of data, and (ii) trade secrets³.

First, the TRIPs Agreement contains a provision relating to the protection of compilations of data. Article 10.2 relating to "compilations of data" provides that:
"Compilations of data or other material, whether 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. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself".

Under this provision, protection of compilations of data is mandatory.

But the second sentence of this Article provides that the protection of compilation of data cannot extend to the data itself. The construction of this second sentence is still open, particularly as to whether it prohibits any protection of mere data by an IP right.

Second, the TRIPs Agreement contains a provision dedicated to trade secrets that can apply to data. Article 39 (section 7: Protection of undisclosed information) provides that:

"1. In the course of ensuring effective protection against unfair competition as provided in Article 10bis of the Paris Convention (1967), Members shall protect undisclosed information in accordance with paragraph 2 and data submitted to governments or governmental agencies in accordance with paragraph 3.

2. Natural and legal persons shall have the possibility of preventing information lawfully within their control from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices so long as such information:

³https://www.wto.org/english/docs_e/legal_e/27-trips.pdf

- (a) *is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;*
 - (b) *has commercial value because it is secret; and*
 - (c) *has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.*
3. *Members, when requiring, as a condition of approving the marketing of pharmaceutical or of agricultural chemical products, which utilize new chemical entities, the submission of undisclosed test or other data, the origination of which involves a considerable effort, shall protect such data against unfair commercial use. In addition, Members shall protect such data against disclosure, except where necessary to protect the public, or unless steps are taken to ensure that the data are protected against unfair commercial use”.*

Scope of this Study Question

- 8) This study question addresses the issue of rights in data, in particular IP rights in data.
- 9) It examines the extent to which data already enjoys protection under current IP and other laws, as well as any gaps or overlaps, such as those that may exist with regard to databases under copyright law and trade secret law.
- 10) It addresses whether there is a need for a new *sui generis* right in certain kinds of data, or whether current laws and contractual agreements are sufficient; and the potential right holder, object and scope of protection of any new right.
- 11) It also addresses whether such a right in data might undercut the existing system of intellectual property rights, unduly restrict the public domain and fundamental rights, distort competition, and hinder scientific research.
- 12) This study question does not address legal issues of privacy and personal data, *i.e.* information relating to an identified or identifiable natural person. Legislation and policy issues relating to personal data protection should not be taken into consideration to answer this questionnaire.
- 13) This study question raises health data as one example of a data-intensive industry where the issues of rights and access to data are important. This question does not address any issues relating to procedures for obtaining legal approvals for products or procedures, such as pharmaceutical approvals and the like.
- 14) This study question also raises the topic of Public Sector Information (PSI) as another large source of data to which access may be desired by certain parties for commercial or other purposes.

Definitions

- 15) In the context of this study, the term “Data” means any mere information (individual item) of any kind, not aggregated and not arranged in a systematic or methodical way, that is recorded and stored by electronic or other means.

- 16) The terms “Database” and “Dataset” mean a collection of independent works, data arranged in a systematic or methodical way and individually accessible by electronic or other means⁴.

Previous work of AIPPI

17) AIPPI has adopted three resolutions in the past:

- Resolution on database protection at national and international level (Q182) in 2004⁵.

This resolution stated, *inter alia*:

“AIPPI recommends that all countries provide for the protection of databases which require substantial investment by means of the sui generis right or other proprietary right subject to the provisions set out below” (first paragraph of the resolution).

But the recital stated that: *“Such proprietary right should not extend to the information and data contained in the database”* (recital b).

One aim of this new study question is to determine whether the recommendation of AIPPI to protect databases has been followed by national/regional legislations.

Another aim is to determine whether the protection of data should now be extended to mere data.

Other relevant prior work includes:

- AIPPI’s Resolution on “Exceptions to copyright protection and the permitted uses of copyright works in the hi-tech and digital sectors” (Q216 and Q216B) in 2010⁶.
- AIPPI’s Resolution on “Copyright in artificially generated works” in 2019⁷.

18) While these Resolutions address related issues, they do not directly address the broader issue of protection of data and data ownership. Thus, it is considered appropriate to continue this avenue of study with this study question on rights in data.

Discussion

19) Under national and regional legislation, *mere data* has traditionally fallen outside the scope of IP protection.

⁴ This definition is set out in the EU directive 96/9/EC of 11 March 1996 on the legal protection of databases, Art. 1. This definition has been adopted by the AIPPI resolution on Database Protection at National and International Level (Q 182, point 3).

⁵<https://aippi.org/wp-content/uploads/committees/182/RS182English.pdf>

⁶<https://aippi.org/wp-content/uploads/committees/216/RS216English.pdf> and <https://aippi.org/wp-content/uploads/committees/216B/RS216BEnglish.pdf>

⁷https://aippi.org/wpcontent/uploads/2019/10/Resolution_Copyright_in_artificially_generated_works_English1.pdf

20) But in many jurisdictions, *databases* can be protected by copyright, if they fulfil the general requirements for protection under copyright, and/or trade secret or unfair competition laws.

21) Furthermore, some jurisdictions have adopted other regimes to provide some measure of protection to investments in creating, collecting and organising databases.

These regimes use different instruments to pursue different goals, with different economic rationales.

For instance, in the EU, collections of data that are the product of substantial investment have been protected by a *sui generis* database right since 1996, with little demonstrable success in incentivising investment⁸.

In 2017, the European Commission floated the idea of creating an EU-wide 'data producer's right' that would protect industrial data, but no follow up initiatives were advanced.

22) The aim of this Study Question is to decide whether such specific regimes providing protection of databases are desirable and sufficient, and whether mere data should also be protected in certain circumstances.

23) The process of creating and using mere data and of databases can be divided into three steps.

1. CREATION/PRODUCTION OF MERE DATA

24) Data exists when it is (i) recorded and (ii) stored in a device.

Many industries have been highlighted as big data intensive industries. This is, for instance, the case for automotive and healthcare industries.

Example 1: the automotive industry (data generated by sensors).

The automotive sector may be trending toward a situation where any device consists of two assets: the physical equipment itself and the data generated from its operation.

Automotive manufacturers assemble each vehicle from many components acquired from subcontractors. These components include data sensors, recorders, and communication units.

⁸ The EU Commission conducted in 2017 and 2018 the evaluation of this Directive: <https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection>

Who is the owner of the data produced during vehicle operation? The subcontractor, the automotive manufacturer, the owner of the vehicle (e.g. a leasor or employer), or the end user? ⁹

Similar issues arise, for example, with collection of vehicular traffic data and weather data.

25) This raises the question of the *rights on mere data*.

In most of the jurisdictions, mere data seems not to be protected by copyright or another IPR. However, the data from millions of cars, taken together, has immense value.

Control of and access to mere data are typically controlled by contract. In most cases, such data is subject to secrecy obligations, thereby preventing access to that data by competitors and the public. Particularly in industries with a small number of large players, this may result in an “information monopoly” that provides significant economic advantage.

The question then is to consider whether control, access and use of mere data should be subject to a specific legal regime, e.g. a new *sui generis* right, with specific prerogatives for the owner and specific exceptions and limitations to the monopoly.

2. CREATION OF DATABASE

26) The structuring and annotation of mere data is crucial, for instance for machine learning.

The transformation of mere data to a suitable form for training an AI model usually requires skill and effort, and can be extremely time consuming and expensive¹⁰.

27) This raises two issues: (i) the *protection* of the resulting database by IP rights, and (ii) the *access to and use* of the mere data to be stored in the database.

28) (i) Database protection

A database can be subject to IP rights if it meets the corresponding legal requirements.

Indeed, it is generally considered that a database which, by reason of the selection or arrangement of its contents, constitutes the author's own intellectual creation shall be protected as such by copyright.

Databases can be protected by others means. For instance, in the EU, a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of its contents can be protected by a *sui generis* right. The data producer can prevent extraction and/or re-

⁹J. Ritter and A. Mayer, Regulating Data as Property: A New Construct for Moving Forward, Duke Law & Technology Review, vol. 16, n°1, p. 221 s.
<https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1320&context=dltr>

¹⁰Machine learning is not the only expression of AI, but it is currently the most developed and efficient one. Machine learning is based on the use of a data set to train an AI model. Consequently, the availability of appropriate training data is critical.

utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database¹¹.

A database can also be protected as trade secret or by unfair competition law if it fulfils the corresponding legal requirements.

The issue is to determine whether the current regimes are satisfactory or if some modifications and harmonization are desirable in light of the rapid developments in this field.

29) (ii) Access to mere data.

Mere data used by the database producer to create the database may have been produced by *third parties*.

Example 2: Works generated by AI (machine learning)

Existing works (e.g. paintings) are selected to be integrated in a database.

This database is used by an AI system to create an artificially generated work¹².

Source works can be protected by IP right, e.g. copyright. Indeed, when a literary, audio-visual or musical work is integrated in a database, for instance to create AI training data, its copyright protection may extend to such use.

Example 3: Applications based on social media

Many applications are based on statistical studies of information available on social media.

An AI system may be used to examine, for example, the huge number of messages (texts, photos, etc.) posted every day on social media and detect certain relationships or trends in order to provide statistics in very diverse domains (health, marketing, etc.).

In these examples, the crucial issue is the lawful access to data.

Is there a need for clarification regarding the use of mere data available on internet? Should specific exceptions to IP rights be implemented, for instance for data mining?¹³

¹¹Article 7.1 of the Directive 96/9 of 11 March 1996 on the legal protection of databases.

¹²See the resolution and study guidelines on Artificially generated works (2019).

https://aippi.org/wp-content/uploads/2019/02/Study-Guidelines_Copyright_Copyright-in-artificially-generated-works_22January2019_NEW.docx.pdf

¹³For instance, the EU Directive 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market, provides an exception to copyright for text and data mining for the purposes of scientific research (Article 3 and 4).

- 30) In other circumstances, third parties may wish to obtain access to existing databases as a whole.

Example 4: Health data (access to a database)

Research in the area of health, and valuation of new medical products and processes, may be greatly facilitated by access to existing databases containing health-relating information. However, such data are confidential and highly valuable information. They may be protected by an IP right (copyright, *sui generis* right) or by trade secret. A researcher may wish to obtain access to such databases for furtherance of health-related studies or initiatives.

The potential desire of third parties to access existing databases is far from limited to this one example. Other examples include browsing and shopping histories, location data, music and video preferences, etc.

3. CREATIONS MADE USING A DATABASE

- 31) When a database is used, for instance by an AI system, the result (invention, work, etc.) may be protected by an IP right (patent, copyright, related right, etc.).

The issue is to determine whether or not the owner of the mere data and/or the database could claim any right on the result of the AI process.

For instance, the resolution on Copyright in artificially generated works (2019) states that:

“AI generated works should only be eligible for protection by Copyright if there is human intervention in the creation of the work and provided that the other conditions for protection are met. AI generated works should not be protected by Copyright without human intervention.

This principle is considered to apply to the Working Example as follows: (...)

“the AI generated work should be eligible for Copyright protection where there are human data selection criteria for the input into the AI”.

If the “human data selection criteria” confers originality to the final AI generated works, the selector could possibly claim copyright on the final work.

You are invited to submit a Report addressing the questions below. Please refer to the 'Protocol for the preparation of Reports'.

Questions

I. Current law and practice

Please answer all questions in Part I on the basis of your Group's current law.

A. Protection of mere data

- 1) Can mere data (in general or some specific mere data) be subject to a property right / IP right? If yes, please answer the following sub-questions:
 - a) What type of property right / IP right would this be?
 - b) What are the requirements for such protection?
 - c) Who is the owner of this property right?
 - d) What acts are prohibited for third parties to avoid infringement?
 - e) Is this right marketable? If so, are specific rules in contract law applicable?
 - f) Does your legislation/case law contain specific exceptions to this protection (e.g. access right for data mining, scientific research, etc.)?

- 2) Is mere data protected by provisions other than a property right / IP right? If yes, please answer the following sub-questions:
 - a) What type of protection is available?
 - b) What are the requirements for such protection?
 - c) Who is the person entitled ?
 - d) What acts are prohibited for third parties to avoid infringement?
 - e) Are mere data marketable? If so, are specific rules in contract law applicable?
 - f) Does your legislation/case law contain specific exceptions to this protection (e.g. access right for data mining, scientist research, etc.)?

B. Protection of databases

- 3) Can a database be subject to a property right / IP right? If yes, please answer the following sub-questions:
 - a) What type of property right / IP right would this be?
 - b) What are the requirements for such protection?

- c) Who is the owner of this property right?
 - d) What acts are prohibited for third parties to avoid infringement?
 - e) Does your legislation/case law contain specific exceptions to this protection (e.g. access right for data mining, scientist research, etc.)?
- 4) Are databases protected by any provision other than a property right / IP right? If yes, please answer the following sub-questions:
- a) What type of protection is available?
 - b) What are the requirements for such protection?
 - c) Who is the person or entity entitled to this protection?
 - d) What acts are prohibited for third parties to avoid infringement?
 - e) Does your legislation/case law contain specific exceptions to this protection (e.g. access right for data mining, scientist research, etc.)?

C. Public Sector Information (PSI)

- 5) Does your legislation contain regulation/case law regarding PSI? if YES, please explain.
- 6) Is there a right to access such PSI?

D. Health data

- 7) Does your legislation contain regulation/case law regarding health data? If YES, please explain.
- 8) Is there a right to access such information?

II. Policy considerations and proposals for improvements of your Group's current law

Could any of the following aspects of your Group's current law or practice relating to rights in data be improved? If YES, please explain and answer each of the sub-questions.

- 9) Protection of mere data?
 - a) Requirements for such protection(s)?

- b) Ownership of the right(s)?
 - c) Scope of the protection?
 - d) Exceptions to this protection?
- 10) Protection of databases?
- a) Requirements for such protection(s)?
 - b) Ownership of the right(s)?
 - c) Scope of the protection?
 - d) Exceptions to this protection?
- 11) Rules on contract law, e.g., prohibition of contractual override, etc.?
- 12) Are there any other policy considerations and/or proposals for improvement to your Group's current law falling within the scope of this Study Question?

III. Proposals for harmonisation

Please consult with relevant in-house / industry members of your Group in responding to Part III.

- 13) In your opinion, should the protection of mere data and/or database be harmonized? For what reasons?

If YES, please respond to the following questions without regard to your Group's current law or practice.

Even if NO, please address the following questions to the extent your Group considers your Group's current law or practice could be improved.

Protection of mere data

- 14) Should mere data be subject to a specific protection, e.g. an IP right or other type of right?
- 15) If yes, what should be the requirements for such protection?

- 16) Who should be the owner of this right / IP right?
- 17) What acts should be prohibited to third parties to avoid infringement?
- 18) Which exceptions, if any, should apply to this protection (e.g. access right for data mining, etc.)?
- 19) What role should contract law play (e.g., prohibition of contractual override)?

Protection of databases

- 20) Should databases be subject to a specific protection, e.g. an IP right or other type of right?
- 21) If yes, what should be the requirements for such protection?
- 22) Who should be the owner of this right / IP right?
- 23) What acts should be prohibited to third parties to avoid infringement?
- 24) Which exception should apply to this protection (e.g. access right for data mining, etc.)?
- 25) What role should contract law play (e.g. prohibition of contractual override)?

Specific regimes

- 26) Should Public Sector Information (PSI) be subject to a specific regime, e.g. regarding the control and access to such data/databases?
If YES, please explain the desirable regime.
- 27) Should health data be subject to a specific regime, e.g. regarding the control and access to such data/databases?
If YES, please explain the desirable regime.

General

- 28) Please comment on any additional issues concerning any additional aspect of Rights in Data you consider relevant to this Study Question.
- 29) Please indicate which industry sector views provided by in-house counsel are included in your Group's answers to Part III.