Landscape Study of Potentially Essential Patents Disclosed to ETSI

European
Commission

Landscape Study of Potentially Essential Patents Disclosed to ETSI

A study carried out in the context of the EC 'Pilot Study for Essentiality Assessment of Standard Essential Patents' project

Authors: Rudi Bekkers, Emilio Raiteri, Arianna Martinelli, Elena M. Tur

Editor: Nikolaus Thumm

2020

Contact information
European Commission, Joint Research Centre
Address: Edificio Expo c/Inca Garcilaso 3, 41092 Sevilla
Email: JRC-LIST-B6-SECRETARIAT@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC121411

PDF ISBN 978-92-76-20786-3 doi:10.2760/313626

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Rudi Bekkers, Emilio Raiteri, Arianna Martinelli, Elena M. Tur. *Landscape Study of Potentially Essential Patents Disclosed to ETSI*. Nikolaus Thumm (ed.). Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-20786-3, doi:10.2760/313626, JRC121411.

Contents

Executive Summary

In November 2017, the European Commission issued a Communication 'Setting out the EU approach to Standard Essential Patents'[1] in which it calls for increasing transparency on SEP[2] exposure. In 2020, the European Commission published a commissioned report[3] that aims to evaluate the technical and institutional feasibility of performing large-scale essentiality checks of patents, titled 'Pilot project for essentiality checks of Standard Essential Patents'.

This study is a companion to the above report. Since patents disclosed at Standards Developing Organisations (SDOs) may be a starting point of such essentiality analyses, this study aims to (1) provide a patent landscape analysis of SDO disclosed patents (and what this implies for their use as input to an essentiality assessment mechanism) and (2) analyse whether SDO disclosed patents differ from comparable other patents in quality (both technical merit and economic value). This study focuses on patents disclosed to the European Telecommunications Standards Institute (ETSI), This SDO is not only very important in the field of telecommunications, but also maintains one of the best databases of such disclosures to date.

The findings of our study are as follows:

1. It is a non-trivial task to identify patents from an SDO disclosure database and clean/ harmonize/select/de-duplicate/transform that data into information to be used for a given purpose, such as input for a process of essentiality assessment. Also important is a good understanding of both the SDO's IPR policy, the related procedures, and database aspects. This study provides guidance for people than want to analyse SDO databases.
2. Disclosed patents at SDOs are self-reported by companies on the basis of their belief that these patents "may be or may become essential". This has important implication for what this data means, for what purpose it is fit to use and for which purpose it is not. Please refer to Figure 14, on page 28, for an overview of the inherent limitations of this type of data.
3. ETSI-disclosed patents can be used as a starting point for an essentiality assessment procedure. For a full assessment (e.g. not sampling). a total of 8,320 to 11,262 (granted) patents would need to be assessed, dependent on the geographical scope of the SEPs exposure to be examined. Note that this is only for ETSI disclosures, and that, on a weekly basis, new disclosures are added.
4. In a given patent family, some patents may be actually essentials, while others may not. We strongly recommend taking the ETSI family definition as a starting point, as it is the best way to identify family members of disclosed patents that may indeed also be essential. Other family definitions (e.g. DOCDB and INPADOC) are problematic in that respect.
5. There is strong upwards trend in the number of new patent families being disclosed. Of all current 25,072 families in the ETSI disclosure database, 37% were added in just the last two years.
6. We observe considerable fragmentation in the distribution of companies – or organisations – that disclosed these patents and observe that the distribution is also very skewed. Such distributional characteristics have to be taken into account should any essentiality assessment scheme consider the use of sampling.
7. We observe big shifts over time in terms of the home country of firms disclosing patents. Especially remarkable is the recent increase in shares of disclosed patent families from Chinese firms and, to a lesser degree, from South Korean firms, at the expense of European and US firms.
8. While it would be desirable to break ETSI disclosures related to cellular standards up into, for instance, technology generations (2G, 3G, 4G and 5G), this is currently not possible in a reliable way. This is, among other things, related to the way 3GPP technical specifications series (which are the input to ETSI standards and specification) are structured.
9. We find that disclosed SEPs score higher than comparable, non-disclosed patents on common proxies associated with patent quality, in terms of technical merit of patents, as well as economic value of patents. In this study, we offer an interpretation of that result.

[1] European Commission (2017).
[2] SEP: Standard Essential Patent.
[3] Bekkers et al. (2020).

Acknowledgements

We like to thank the European Telecommunications Standards Institute (ETSI) for support and input. We especially want to thank Christian Loyau for feedback, and Philippe Cavé for spending time with us at the institute's premises in Sofia Antipolis to ensure proper retrieving and processing of ETSI disclosure data. We also want to thank Serge Raes, Vice-Chair of the ETSI IPR Special Committee, for extensive feedback and suggestions.

We also want to thank the European Patent Office (EPO) for input and feedback; especially Ged Owens, Franco Cordera, Yann Ménière, and Michael Fröhlich.

Finally, we also like to express our gratitude to the members of this project's Advisory Board of the 'Pilot project for essentiality checks of Standard Essential Patents' (of which this study is a part): Christian Loyau, Edmund Mangold, Yann Ménière, Ruud Peters, Serge Raes, Mats Sågfors. All the views expressed in the report are those of the authors only and are not necessarily endorsed by the advisory board.

Commissioning body

This study was initiated and financed by the European Commission Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW). It was commissioned by the European Commission Joint Research Centre (JRC), acting in close collaboration with the European Commission DG GROW and the European Commission Directorate General for Communications Networks, Content and Technology (DG CONNECT).

The supervisory board of the European Commission included:

- Dr. Nikolaus Thumm (EC - JRC)
- Ms. Anne von Zukowski (EC - DG GROW)
- Mr. Emilio Davila Gonzalez (EC - DG CONNECT)

1 Introduction

In November 2017, the European Commission issued a Communication 'Setting out the EU approach to Standard Essential Patents' (European Commission, 2017) in which it calls for increasing transparency on SEPs exposure. In March 2020, the European Commission published a commissioned report that aims to evaluate the technical and institutional feasibility of performing large-scale essentiality checks of patents, titled 'Pilot project for essentiality checks of Standard Essential Patents' (Bekkers et al., 2020).

This study is a companion to the above report. The main goals of this study are:

- To provide a patent landscape analysis of SDO disclosed patents (and what this implies for their use as input to an essentiality assessment mechanism)
- To analyse whether SDO disclosed patents differ from comparable other patents in quality (both technical merit and economic value).

Before we go into detail, it is important to stress that the starting point for this study is data that is self-reported by companies on the basis of their belief that these patents **'may be or may become essential to an ETSI standard'** (see Section 2.1). This has important implication for what this data means, for what purpose it is fit to use and for which purpose it is not. Please see Figure 14, on page 28, for more details.

1.1 Approach and challenges

Over the last decade, a lot of studies have analysed the lists of patents that are disclosed as essential or potentially essential, and many of them also investigated features of these patents such as forward citations (Bekkers et al., 2017). To learn how disclosed SEPs possibly differ from other patents, one needs to compare these with patents that are not disclosed but otherwise comparable. There are several challenges here. This chapter describes the methodology we adopted to tackle these challenges and the results of an extensive comparison in terms of patent quality measures between disclosed SEPs and a group of suitable control patents.

The first challenge one needs to deal with for this type of study is to identify a reliable source for disclosure data, a source that would allow to identify the universe of disclosed patents for a specific standard and to clearly identify all the members of the patent family of the disclosed patents. These are necessary requirements for reducing concerns of confounding effects in the analysis. Using a data source that is not complete or does not permit to reconstruct the full patent family of a disclosed patent would expose the analysis to the risk of including patents that have been disclosed as SEPs or members of a family disclosed as SEP. To ensure our analysis met the above requirements, we decided to focus on patents disclosed to a single Standard setting Organization (SSO): the European Telecommunication Standardization Institute (ETSI). ETSI maintains a public and complete database of patents that are disclosed by the patent owners as potentially essential to an ETSI standard. The information available in the ETSI database also allows to reconstruct the patent family of disclosed patents and to match the disclosed SEPs to other data sources to recover additional bibliographic information.

Second, although the ETSI database is public, the data collection and processing require considerable attention and careful decision making. We aimed at creating a list of all the unique patents and patent applications disclosed to ETSI and to identify all the members of the family of the disclosed patents. To this end, we decided to follow the definition of patent family used by ETSI itself, which includes all the documents having at least one priority in common, including the priority document(s) themselves. In addition, we also decided to consider both granted patents and patent applications in our data gathering and identification process. For this reason, unless in the cases in which we clearly indicate otherwise, whenever we speak of 'patents' in this study, we mean both patent and patent applications (similar to the ETSI IPR policy).[4]

[4] As we will explain later, some parties also disclosed USPTO provisional patent applications to ETSI. These disclosed patents are not harmonized by ETSI, neither part of the PATSTAT database, and not included in our analysis. Yet, in many cases, if this provisional patent application was the basis of a later real patent application, the patent owner has disclosed that application and thus would be in our analysis.

Third, we need to gather additional and up-to-date information about the ETSI disclosed patents. To do so, we matched the list of disclosed SEPs that we recovered from ETSI with patents in EPO's Worldwide Patent Statistical Database PATSTAT (Version 2018b). Of particular importance is the identification and construction of the patent quality measures that are used for the focal analysis.

Finally, the quality comparison between disclosed SEPs and 'comparable' patents not disclosed as SEP demanded the creation of an appropriate control set. To facilitate this task, we decided to consider exclusively patents granted by the USPTO or the EPO, clearly both for the disclosed SEPs and the control patents. Then we apply exact matching technique to increase the degree of similarity of between the two groups, and, given the general abundance of potential control patents for each disclosed SEP, we randomly select up to five control patents for each of the treated unit.

This study is structured as follow: Chapter 2 discusses the data collection and processing; Chapter 3 describes the matching with the PATSTAT database, the construction of our variables of interests, and present descriptive statistics about the patent actually disclosed to ETSI; Chapter 4 describes the creation of the control group, the method of analysis and presents the results. Chapter 5 concludes. Table 1 reports a brief description of the content, data, and methodology used in each in each of the sections.

Table 1. Overview of chapters: Content, data, and methodology

Chapter and main topics	Data	Method	Outcome
2.Data collection: - Identification of disclosed SEPs - Identification of the ETSI patent family	ETSI on-line disclosure database EPO-PATSTAT Database (Autumn, 2018)	Database matching based on application/publication number or family identifier	Identification of 25,072 disclosed (basis) patents and their related ETSI family
3. Disclosed SEP landscape: - Construction of relevant patent variables (patent level) - Descriptive statistics	25,072 disclosed (basis) patents as identified in Chapter 2	Gathering of patent level information on the sample of 25,072 disclosed (basis) patents	Patent-level information dataset Description of the characteristics of disclosed (basis) patents along several dimensions: timing, technology, ownership, patent family size, claims, and backward and forward citations
4. Quality assessment: - Identification of the disclosed SEP - Construction of the control set - Quality assessment	Set 1: 4,607 granted disclosed EPO patents and 19,477 matched control patents Set 2: 12,832 granted disclosed USPTO patents and 56,100 matched control patents EPO-PATSTAT Database (Autumn, 2018) OECD Quality Database (Version 2019)	Starting point is the 25,072 patent families, as identified in Chapter 2. Focal patents are granted EPO (USPTO) patent family members of these families Construction of the control group: exact matching on patent-level characteristics	Quality comparison between disclosed SEPs and patents not disclosed Regression analysis to account for quality differences between disclosed and control patents

1.2 Terminology

In terms of terminology, we stay close to the way terms are defined by ETSI.[5]In particular:

- By 'patent' we refer to patents, utility models, and applications therefor.
- By 'basis patent', we mean a patent disclosed to ETSI as potentially essential, including its patent family (according the ETSI patent family definition).[6]
- By 'Patent family according to the ETSI patent definition', we mean all the patent documents having at least one priority in common, including the priority document(s) themselves.
- Following the terminology in the ETSI database, by disclosing 'company' or 'firm', we refer to any organization that discloses a patent, including public and private research organisations, associations, and universities.
- In the field of essential patents, both the word 'disclose / disclosure' and 'declare / declaration' is used when a party informs an SDO of a patent that may be essential. Also in ETSI's documents, both terms are used. In the context of this study, the terms are exchangeable. For reasons of consistency, we use the word 'disclose' and 'disclosure', unless when we refer to the specific wording used in sources.

[5] For the ETSI definitions, see Clause 15 of the ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

[6] The term 'basis' is used by ETSI itself in its IPR database; see ETSI (2011).

2 Data collection

2.1 The ETSI IPR disclosure database

The European Telecommunication Standardization Institute (ETSI) maintains a public database of patents (as well as patent applications and business models[7]) that are disclosed by their owners[8] in the belief that these patents may be or may become essential to an ETSI standard.[9] As a result, this database contains patents that are potentially essential. In the disclosure form, owners can also indicate that they are willing to commit to license these essential IPR at FRAND terms and conditions, to the extent that these patents actually become essential[10] (which they almost invariable do). Only the declarant can modify its own declaration[11]; ETSI is not entitled to modify a declaration.

The purpose of this database (and the related disclosure) is, in short, that ETSI knows what patents are potentially essential, and knows for which patents a license commitment is provided. Should a license commitment not be provided for a potentially essential patent, then ETSI can avoid adopting a standard that would require the use of that patent.[12]

It is important once more to stress that the patents in this database are potentially essentially, and not necessarily actually essential.

2.2 Patent families and the ETSI disclosure database

It is common that an applicant files for patents on the same invention in multiple jurisdictions. An ETSI member may (voluntary) disclose all or a multiple of these patents, or just disclose one of them. Yet, to ensure that a FRAND licensing commitment still covers all these patents, ETSI specifies that the licensing commitment is deemed to be made for all existing and future members of the patent family of that patent family.[13] An important aspect here is to define what a patent family exactly means in then ETSI context. But before doing so, we first shortly discuss two other common patent family definitions, known as the DOCDB family and the INPADOC family, used by the European Patent Office, and others.

---

7 The ETSI IPR policy covers IPR, which are defined as “any intellectual property right conferred by statute law including applications therefor other than trademarks. For the avoidance of doubt rights relating to get-up, confidential information, trade secrets or the like are excluded from the definition of IPR.” Furthermore, a “Patent family” in ETSI is defined as “all the documents having at least one priority in common, including the priority document(s) themselves. For the avoidance of doubt, “documents” refers to patents, utility models, and applications therefor.” ETSI Intellectual Property Rights Policy, Clause 15, in Annex 6 of ETSI (2018a). In the context of this study, patents, patent applications and business models are the most relevant (there are very few, if any other IPR disclosed to ETSI). In this study, when we refer to ‘patents’, we mean patents, patent applications and business models.

8 While at ETSI a party can also disclose a patent owner by another party (i.e. by mentioning it at a meeting, or informing ETSI by letter), the disclosures as we find them in the ETSI IPR database are also licensing commitments and are submitted by the patent owner.

9 The precise language in the ETSI ‘Information Statement and Licensing Declaration” (ISDL) via which such disclosures are submitted is: *“In accordance with Clause 4.1 of the ETSI IPR Policy the Declarant and/or its AFFILIATES hereby informs ETSI that it is the Declarant’s and/or its AFFILIATES’ present belief that the IPR(s) disclosed in the attached IPR Information Statement Annex may be or may become ESSENTIAL in relation to at least the ETSI Work Item(s), STANDARD(S) and/or TECHNICAL SPECIFICATION(S) identified in the attached IPR Information Statement Annex.”* The referred Clause §4.1 reads as follows: *“Subject to Clause 4.2 below, each MEMBER shall use its reasonable endeavours, in particular during the development of a STANDARD or TECHNICAL SPECIFICATION where it participates, to inform ETSI of ESSENTIAL IPRs in a timely fashion. In particular, a MEMBER submitting a technical proposal for a STANDARD or TECHNICAL SPECIFICATION shall, on a bona fide basis, draw the attention of ETSI to any of that MEMBER's IPR which might be ESSENTIAL if that proposal is adopted.”*. ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

10 The patent owner may also make its commitment subject to the condition that those who seek licensees agree to reciprocate. See Clause 6.1, ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

11 Note, however that while updates are allowed, a licensing commitment entered into is irrevocable. See Clause 6.1, ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

12 The overall purpose of the ETSI IPR policy is as follows: “to reduce the risk to ETSI, MEMBERS, and others applying ETSI STANDARDS and TECHNICAL SPECIFICATIONS, that investment in the preparation, adoption and application of STANDARDS could be wasted as a result of an ESSENTIAL IPR for a STANDARD or TECHNICAL SPECIFICATION being unavailable.”. Clause 1, ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

13 “The obligations pursuant to Clause 4.1 above are deemed to be fulfilled in respect of all existing and future members of a PATENT FAMILY if ETSI has been informed of a member of this PATENT FAMILY in a timely fashion. Information on other members of this PATENT FAMILY, if any, may be voluntarily provided.” ETSI Intellectual Property Rights Policy, Clause 4.3, in Annex 6 of ETSI (2018a).

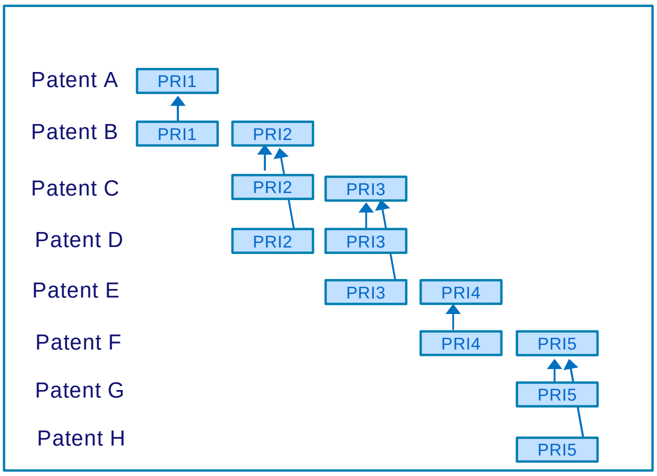
To better understand the various patent family definitions and their implications, we will use an example of eight patents. A key element here are the priority documents that can be listed in patent applications. The presence of a priority document indicates that (some) elements of the application are already part of an earlier application, and that the examiner should use the data of that priority document in order to determine the relevant prior art when examining that patent. This might be the case if an application for the same invention has already been filed abroad,[14] but it is also possible if an application builds upon an element already present in another application filed in the same country, or, in the US, on a provisional patent application.

Now, for our example, assume that our eight patents that have the following priority relationships:

- Patent A lists no priority documents. We will call the invention in this patent Priority 1 (PRI1)
- Patent B lists Patent 1 as a priority. It also adds some novel elements; which we will call Priority 2.
- Patent C lists Patent 2 as a priority. It also adds some novel elements; which we will call Priority 3.
- Patent D lists both Patent 2 and Patent 3 as priorities. It does not add any new elements (for instance, it could be the same patent application text as Patent C but filed in another country).
- Patent E lists Patent 3 as a priority. It also adds some novel elements; which we will call Priority 4.
- Patent F lists Patent 5 as a priority. It also adds some novel elements; which we will call Priority 5.
- Patent G lists Patent 6 as a priority. It does not add any new elements (for instance, it could be the same patent application text as Patent F but filed in another country).
- Patent H lists Patent 6 as a priority. It does not add any new elements (for instance, it could be the same patent application text as Patent F but filed in another country).

Figure 1 now shows our eight patents, where the arrows represent the priorities listed by the patents, and the underlying priorities elements (which consist of one or more patent claims) are shown as PRI1 thru PRI5.

Figure 1 Example of eight patents and their priority relationships



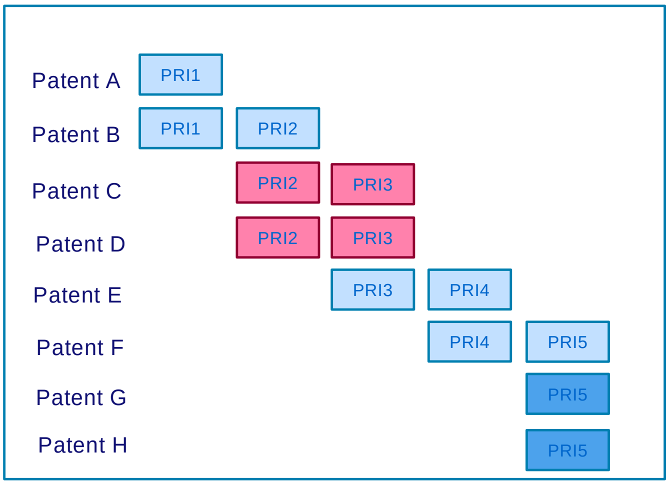
[14] Under the right of priority, established by the Paris Convention for the Protection of Industrial Property of 1883, an applicant that filed for a patent one country can within 12 months, apply for protection for that invention in all the other countries. These later applications will then be regarded as if they had been filed on the same day as the earliest application. Via the so-called PCT route, the applicant has additional possibilities, even after 12 months. See WIPO (2008) for more details (especially §5.20, §5.252).

In a **DOCDB family**, patents are member of the same family if they share *exactly* the same priority documents. Figure 2 illustrates this for the eight patents in our example. Here,

- patents C and D together form an DOCDB family,
- patents G and H together form an DOCDB family,
- all other patents each form their own DOCDB family, with just a single member.

Members of a DOCDB family are very, very similar, which makes this family useful for a number of purposes. However, in the context of ETSI and the objective of ETSI to ensure FRAND commitments for all potentially essential patents, this family has disadvantages. Suppose Priority 2 (PRI2) in our example is the element that gives rise to potential essentiality. If a party would disclose Patent C, then the family would not encompass Patent B, even though it would include the potential essential element too. Furthermore, note that DOCDB families are, by their definition, always mutually exclusive, meaning that a patent can never be member of more than one family.

Figure 2. The six DOCDB families in the set of patents shown in Figure 1. Patents C and D form one family, patents G and H one families, and all other patents are a one-member family.

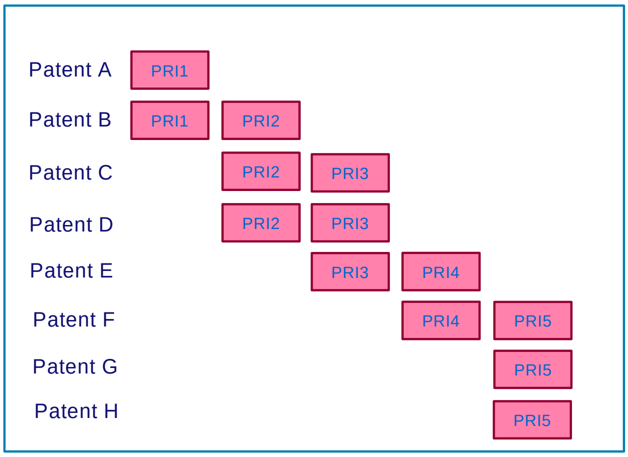


Priorities related to two DOCDB families are highlighted: the family with patent C and D (red) and the family with patent G and H (dark blue). All other patents, not highlighted, are each families by themselves.

In an **INPADOC family**, patents are member of the same family if they share at least one priority with at least one other family member. Figure 3 illustrates this for the eight patents in our example. Here,

- Patents A, B, C, D, E, F, G and H together form an INPADOC family.

Figure 3. The one single INPADOC family in the set of patents shown in Figure 1.



Members of an INPADOC family may be rather different from each other and may actually not directly share any priority (like patents A and H in our example, as well as many other example). For that reason, in the context of ETSI and the objective of ETSI to ensure FRAND commitments for all potentially essential patents, this family has some disadvantages too. Suppose Priority 2 (PRI2) in our example is the element that gives rise to potential essentiality. In that case, patents B, C and D are potentially essential. Patents A, E, F, G and H would not, but would still be part of the family for which a commitment is entered into. Note furthermore that in the INPADOC definition, families are mutually exclusive, meaning that a patent can never be a member of more than one family (just like in DOCDB families).

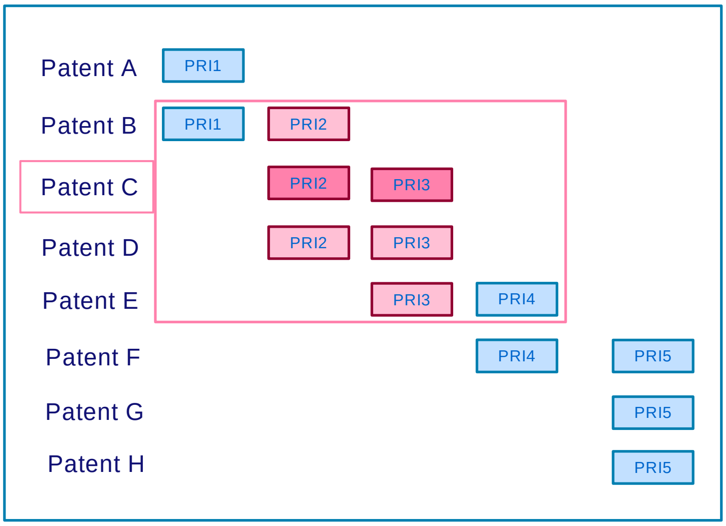
As discussed above, neither the DOCDB or the INPADOC family definition are very suitable for the disclosure and FRAND licensing commitment process at ETSI. This is probably the reason why ETSI provides its own definition of a patent family, which reads as follows: "*all the documents having at least one priority in common, including the priority document(s) themselves*".[15] Using the same example as above, Figure 4 shows how the patent families following the ETSI family definition would look like. Let's focus on patent C and assume that it is a patent disclosed to ETSI. Following the same language as ETSI, we will call the disclosed patent the **basis** patent.[16]

Here, basis patent C is referenced by patent D and E as a priority document. In addition, patent C itself references patent B as a priority document. So, following the ETSI definition, from the perspective of Patent C, Patents B, D and E are family members.

15 ""PATENT FAMILY" shall mean all the documents having at least one priority in common, including the priority document(s) themselves. For the avoidance of doubt, "documents" refers to patents, utility models, and applications therefor." Clause 15 of the ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

16 ETSI (2011).

Figure 4. Illustration of the ETSI family definition of the disclosed ('basis') patent C.



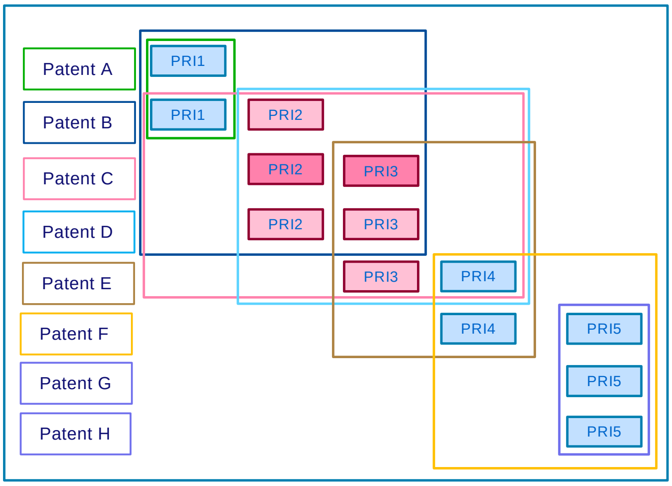
Only the priorities related to family of disclosed ('basis') patent C are highlighted. Red indicates the priorities of patent C itself, pink indicates the priorities patent C directly shares with other patents.

Family member relationships for the other patents are as follows:

- From the perspective of Patent A, Patent B is a family member,
- From the perspective of Patent B, Patents A, C and D are family members,
- From the perspective of Patent D, Patents B, C and E are family members,
- From the perspective of Patent E, Patents C, D and F are family members,
- From the perspective of Patent F, Patents E, G and H are family members,
- From the perspective of Patent G, Patents F and H are family members,
- From the perspective of Patent H, Patents F and G are a family member.

This full set of relationships is shown in Figure 5.

Figure 5. The seven families following the ETSI family definition in the set of patents shown in Figure 1. Note that these are ego-families (as seem from one specific patent).



From the perspective of the ETSI disclosure obligation, the ETSI patent family definition has clear advantages over DOCDB and INPADOC family definitions. The ETSI family definition ensures that all patents that may also include the claims that give rise to essentiality are indeed part of the family, but that patents that do not include the claims that give rise to essentiality (because they do not share any direct priority with the disclosed patent) are not part of the family.[17]

Note that patent families (following the ETSI family definition), by their mere definition, also have two characteristics that are notably different from DOCDB, INPADOC, and many other patent family definitions:

- they are ego-families: they are only valid from the perspective of the disclosed ('basis') patent;
- they are non-exclusive.

This means that one should exercise caution with using these families in a cumulative way in order to determine the scope of the disclosed patent portfolio for a given company, for instance. Since patent families following the ETSI family definition can overlap, there may be (and there will likely be) duplication between those families. Also, while it is certainly possible to do a citation analysis for an ETSI family as a whole, one should exhibit caution with 'adding up' the results of the different families of a given company, as there will be citation duplication between these observations.[18]

Therefore, in this study, we will use the ETSI family definition where appropriate (e.g., in terms of disclosure of potential SEPs but use other family definitions where necessary.

Summarizing: the set of patents shown in Figure 1 encompasses six DOCDB families, one INPADOC family, and seven families following the ETSI family definition (the latter being ego families, and very different in composition from the six DOCDB families).

2.3 Retrieving and preparing data from the ETSI disclosure database

The ETSI disclosure database is publicly available in two different forms:

1. The ETSI Special Report SR 000 314, which is snapshot of the full ETSI database that is made available twice a year. By the time of writing this document, the most recent version was version V2.26.2, issued in November 2019.[19]
2. The ETSI on-line disclosure database, available at ipr.etsi.org. This form delivers the most up-to-date data. It has two different parts to it:
    a) 'declaration search', where specific disclosures can be found, including the original disclosure itself (usually in facsimile format) as well as information on all the basis patents provided by the company in that disclosure and the ETSI family members of those patents (as automatically determined by the ETSI secretariat in collaboration with the EPO). The full history of a disclosure is made available as well. Note that disclosures often exist of numerous (numbered) disclosures, and that disclosures can exist of multiple basis patents and many family members.
    b) 'Dynamic Reporting', which allows users to create a dataset of selected patents, or the entire disclosure database, on the basis of a number of search criteria.

For this study, we used the 'Dynamic Reporting' tool, as it allows to create a data set for the full disclosure database and is richest in terms of available data.[20] We used this data as it was available as of 31 February 2018.[21]

17 Note that also the ETSI family definition is not *perfect* in the context of disclosure. Should, in our example of disclosed basis patent 3, only PRI2 give rise to essentiality, but not PRI3, then patent 5 would be within the patent family definition but not be essential itself. Having that said, an alternative definition that would not have this disadvantage would need to rely on priority information on the individual claim level, and such information is not available in the patent system. As a consequence, a third party, such as a prospective licensee, would have a lot of trouble finding out for which patent the FRAND commitment is exactly entered into. Given this consideration, the ETSI definition is probably the best possible compromise here, and still considerably more appropriate than the DOCDB or INPADOC definition

18 On strategy to deal with such citations is to count 'distinct citing patent families' instead of counting all citing patents (or patent families).

19 This Special Report is publicly available at [www.etsi.org] using search term, "SR 000 314"

Our full retrieved data set from the Dynamic Reporting tool included 706,594 records of each eleven fields.[22] In Table 2, an example of such a record is given. In this tool, a record is a unique combination of data in its eleven fields, but there may be significant duplication between records (e.g., another record may have exactly the same information in all or some patent fields but different information in the 'ETSI Projects' or 'Standards' field.).

Table 2. Example of record in ETSI 'Dynamic Reporting' output

Field name	Value
ETSI Projects	3GPP NR Rel 15
Standards	None
Companies	Sharp Corporation
Patent Families	|JP20110169318|WO2012JP69672|
Application Number	US201214236541
Publication Numbers	US2014177584 A1|US9801143 B2
Title	TERMINAL, COMMUNICATION SYSTEM, AND COMMUNICATION METHOD
Patent Offices	US (UNITED STATES)
Declarations	ISLD-201802-006
IPR Disclosures	Disclosure number 194
Declaration Date	20/02/2018

There are three main[23] fields that provide information on the disclosed patent in question:

1. The **'Application Number'** field provides information on the number assigned to the patent office to the application of the disclosed patent. If the patent is 'harmonized' by ETSI (see below), it is provided in a format known as EPODOC[24]. If not, the field contains the input as provided by the discloser. In that case, the format of the number is not harmonized, and in many cases, the number is incomplete or even erroneous (e.g., not an application number, but a publication number instead).
2. The **'Publications Number'** field provides one or two document identifiers of publications by a patent office related to that patent. These publications may be publications of a patent application (usually bearing the 'kind codes' A1, A2 or A3 as last characters of the number), publications of a granted patent (usually bearing the 'kind codes' B1 or B2 as last characters of

20 For instance, the data from the Dynamic Reporting tool includes patent family information (given as application numbers of the priority documents) as well as up to two publication numbers of the disclosed patents.

21 The dataset we use was created on 31 February 2018, at the premises of ETSI in Sofia Antipolis, France, with the kind help of the ETSI legal/ICT staff, to ensure the correct retrieval and interpretation of data.

22 In the Dynamic Reporting tool, we selected the following fields: "ETSI Projects", "Standards", "Declaring companies", "Patent family" and "Patent". We only selected patents for which the disclosing company stated it to be 'essential'. (Note that these selection fields are precisely the same as the fields that appear in the output of the tool.) Following advice of the ETSI staff, we did not select "Patent Office" or "Patent holder", as these are known to be problematic if the full database is retrieved. For 'Patent holder', for instance, the system would need to retrieve 'life' data from the EPO on recent ownership, and for many thousands of patents, that could take very long. As export options, again following advice of the ETSI staff, we selected "Export of detailed patent information", "Declaration reference", "Declaration date", and "IPR disclosure", and "Export of basis patent only", and did not select "Export of detailed patent information". Making other choices for downloading the full dataset (e.g. exporting all patents instead of only the base patent) could result in excessive file sizes (>10GB) which the ETSI server will not be able to generate.

23 A fourth and fifth field are the "Title" and the "Patent Offices" fields. The title field (taken by itself) is not useful for identification purposes as it may not be unique (there may be other patents with the same title). For harmonized records, the "patent offices' field does not provide any information that is not already in the other fields. For unharmonized records it may be informative (e.g. in cases where the discloser did not provide country information in other fields), but we do not consider these records, as explained below.

24 EPODOC is a unique number to identify patent applications and is created by the EPO based on the DOCDB application number, application authority and application kind. See EPO, Data Catalog PATSTAT Global, Version 5.12 (for PATSTAT 2018 Autumn Edition), dated 01 October 2018.

the number), or other publications. If the patent is 'harmonized' by ETSI (see below), it is provided in a harmonized format of <country code> <publication number> <kind code>. Note that the numbering convention are different per patent office. For instance, the EPO uses the same number for the publication of a patent application and the associated publication of the patent grant (e.g. EP2741550A1 and EP2741550B1), while the US does not (example: US2014177584A1 and US9801143B2). If the patent is not harmonized by ETSI, the field contains the input as provided by the discloser. In that case, the formatting of the number is not harmonized, and in many cases, the number is incomplete or even erroneous (e.g., not a publication number, but an application number instead).

3. The **"Patent Families"** field provides, for disclosures harmonized by ETSI, application numbers of patents that represent priority documents for the disclosure in question. (Note: this field does not include all family members!). In our retrieved data set, there were up to 27 numbers provided in this field. Sometimes the application number of the disclosed patent itself is also present in this field, sometimes it is not.[25] Numbers in this field are provided in a format known as EPODOC.

These three fields (and the up to 30 values in them) provide three broad routes towards identifying the identity of a patent disclosed to ETSI. It is important to note, however, that patents may be disclosed to ETSI for different 'ETSI Projects', and/or for different "Standards", and/or in different 'Declarations'. This is illustrated in Table 3, which shows all entries in the retrieved data set for which all three fields described above are exactly identical to the example already given in Table 2. Furthermore (but not shown in this example), there might also exist disclosures for the same patents from other declarants (called 'companies' in the ETSI database), for instance by those that acquired a patent that was already previously disclosed at ETSI.

Table 3. Other records concerning the same patent as in Table 1.

ETSI Projects	Standards	Companies	Patent Families	Application Number	Publication Numbers	Patent Offices	Declarations	IPR Disclosures	Declaration Date
3GPP NR Rel 15	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.213	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.213	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.214	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.214	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.331 (15.0.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP NR Rel 15	TS 38.331 (15.0.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201802-006	Disclosure number 194	20/02/2018
3GPP-radio	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 136 213 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 136 213 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 136 331 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 136 331 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 36.213 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 36.213 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 36.331 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-radio	TS 36.331 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	None	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 136 213 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 136 213 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 136 331 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 136 331 v12.5.0	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 36.213 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 36.213 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 36.331 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017
3GPP-Release-12	TS 36.331 (12.5.0)	Sharp Corporation	|JP20110169318|WO2012JP69672|	US201214236541	US2014177584 A1|US9801143 B2	US	ISLD-201707-017	Disclosure number 40	12/05/2017

As our objective for the data gathering/processing phase is to create a list of all unique, disclosed patents to ETSI (see Chapter 1), we combine all records that have precisely the same identification information in three fields discussed above. After this first duplication removing step, we have 48,684 records, considerably less than the 706,594 records before this step. (Note that after this step, there still might be records that have some overlap in patent information, but not fully the same information in all three fields; we will deal with this in a second duplication step which takes place after we matched various fields with PATSTAT).

When performing the above deduplicating step, we always kept the record with the oldest disclosure date, since we primarily interested in the disclosure and its characteristics, not the 'current owner' (determining the actual owner which would, to be reliable, require many more efforts than just taking the latest

[25] For 137,082 (19.4%) of the 706,594 records in our retrieved dataset, there is no patent family information ("NONE"). In the cases where exactly one patent family member is mentioned, it is the same patent as the patent listed in the Application Number field in about half of the cases (48.5%), and a different patent in the other cases.

declarant in the disclosure database). This means that in the resulting data set, 'disclosure date' changes to 'earliest disclosure date', and 'company' to 'Earliest disclosing organization'.[26]

While the above set of 48,684 records is the input to our patent identification process as described below, the eventual number of actually different, unique patents is somewhat lower. Table 4 illustrates why this is the case. In some cases, we see records in the database that refer to exactly the same patents in the Application Number and Publication Numbers fields, yet have different information in the Patent Families fields. The exact reason for this is not known to us, but it might be caused by ETSI constructing family composition for different records at different points in time (and at a later point in time, new family members might appear). Whatever the reason is, in the perspective of our study, such records must clearly be merged, and so we eventually have somewhat less than 48,684 records.

Table 4. Records with the same patent but different family information

Patent families	Application number	Publication numbers
US20010033141 US20010279970P US20010933914	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010933914	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010933971	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010933972	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010933977	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010933977 WO2002US09830	US20010933977	US2002141391 A1 US6707801 B2
US20010279970P US20010934021	US20010933977	US2002141391 A1 US6707801 B2

2.4 Identifying the patents in the ETSI disclosure database

In order to be able to perform the range analyses presented in this study, we attempted to identify all patents in the ETSI database in the EPO's PATSTAT. With over 100 million patent records from 90 patent issuing authorities, covering the mid-19th century up to today, and with two updates every year, PATSTAT has become a point of reference in the field of patent intelligence and statistics. To be able to identify as many patents as possible from our retrieved ETSI database (of late February 2019), we used the latest available PATSTAT version, which was the 2018 Autumn Edition.[27]

We used all three fields in the ETSI database, described in the previous section, which can identify the patent in question. In some cases, a patent could only be identified from one field, in other cases from two or all three. If patents could be identified from more than one field (or by multiple values in that field), we used that information for consistency checks, allowing us to see when our matching instruments in one or more fields was not working well (or not working well for some specific patent offices, numbering ranges, formatting types, etc.).

In principle, the data in the three above fields *should* match if the record was harmonized by ETSI, and *may* be matched if they were not harmonized by ETSI by nevertheless contained values in the same

[26] This step also meant we had to drop the information for the following fields, as they could have different values: ETSI Projects, Standards, Title, Patent Offices, Declarations and IPR Disclosures. However, when necessary, these can be later easily added to our data (taking into account their possible *1:n* relationship).

[27] Detailed information is available in EPO (2018).

formatting as the ETSI harmonization would have generated.[28] (See below on how we prevented that the latter resulted in erroneous outcomes.)

Table 5. Matching statistics

Matching status	**Records**	**Percentage of all records**	**Percentage of matched records**
Matched by Application Number, Publication Number(s) and Patent Family number(s)	28,561	58.7%	78.4%
Matched by Application Number and Publication Number(s)	4,400	9.0%	12.1%
Matched by Publication Number(s) only	1,110	2.3%	3.0%
Matched by Patent Family number(s) only (*)	912	1.9%	2.5%
Matched by Application Number only	782	1.6%	2.1%
Matched by Publication Number(s) and Patent Family number(s)	394	0.8%	1.1%
Matched by Application Number and Patent Family number(s)	255	0.5%	0.7%
Not matched	12,270	25.2%	n/a
Total	48,684	100.0%	100.0%

Note: *(*): see text*

Table 5 shows our matching statistics. In total, 36,414 records were matched, and 33,610 (92,3%) were matched in at least two independent ways (i.e. categories a, b, f and g in the table). For a total of 12,270 records, no match was found at all; in Section 2.4.2 we will further discuss the sources and consequences of that. We used the following matching priority, we used: (1) the first matched Publication Number, (2) the matched Application Number, or (3) the first matched Patent Family Number (which is a priority document, as explained above).

2.4.1 Checking for possible errors in our identification process

We checked whether our identification process performed well in two different ways: consistency checking and erroneous records checking.

In terms of consistency checking, we found that out of the 32,961 records for which we both were able to match Publication Numbers and Application Number (i.e. categories a and b in Table 5), there were 12 cases (0,036%) where these numbers lead to different information. We looked these cases up manually and hand-picked the correct patent, using the declarant and patent title information.

Furthermore, with patents we matched via the Patent Family Number, we followed a slightly different procedure: as the field 'Patent Family Number' contains information priority document, it does not necessarily have to bring us the exact disclosed patent. But it should, in any case, bring us to the same INPADOC family of the disclosed patents (after all, they share a priority document). Out of the 19,210 records where we had both Patent Family Number matching as well as some other matching (i.e. categories a, f, and g in Table 5), we found 35 cases (0,18%) where the INPADOC families were not identical. Also here, we looked these cases up manually and hand-picked the correct patent, using the declarant and patent title information.

In terms of erroneous records checking, we analysed the technology classes (IPC) of all the 36,414 records with matched results for classes we should not expect to be present in a database of patents disclosed as potentially essential for mobile telecommunications. To this end, we selected all patents for which the

[28] We chose this approach not only because it maximized the potential matching rate, but also because the retrieved ETSI database does not *a priori* allow to establish (at full certainty) whether a given record is harmonized or not.

share of ‘G’ and ‘H’ sections (the highest level of hierarchy of the IPC classification) is less than 50%.[29] We found 123 patents (0.34%) that met that criterion, and after inspecting their titles concluded that all of these were erroneous entries. We removed them from our dataset. We did, however, perform further investigation to understand why they occurred, which revealed that 60 of these 123 patents were by one specific Chinese declarant, all disclosed on the same date and all referring to Chinese application numbers from one single year. The other erroneous records showed no distinguishable pattern (different companies, different match fields, different years) and can, we believe, only be attributed to erroneous inputs at the side of the declarants, or non-systematic identification problems in PATSTAT (but as such low occurrence that they are not a reason for concern).

On the basis of the very low number of inconsistent or erroneous cases, we conclude that all our three matching strategies in PATSTAT were performing properly and consider our data to be reliable for the set of matched patents.

2.4.2 Source and consequences of unidentified matches

As can be seen in Table 5, no matches could be found for 12,270 records (52,2% of all probed records). We now discuss the possible sources and consequences of those nun-matches.

In Table 6, we show examples of non-matched records, selected to illustrate different causes. (We only show the fields relevant to possible identification). Firstly, we see examples that are tough to identify because they are incomplete or do not use any harmonized formatting in the Application Number or Publication Numbers fields. With very considerable manual effort, some patents might still be identified (Examples 1 thru 3), but there are also records (Examples 4 thru 10) that are so incomplete that such a task would be virtually impossible.

[29] We do so because it may be the case that a patent relevant for mobile telecommunications is also found by a patent examiner to be relevant for another category, outsize the G and H sections. A patent that is not present in the G or H section at all, however, is very likely not relevant for mobile telecommunications.

Table 6. Selected examples of non-matched records

Example	Patent Families	Application Number	Publicatio n Numbers	Title	Patent Offices
1	None	942	None	Encryption system for digital cellular communications	VE (Venezuela)
2	None	1031	None	Pop-subscription based mobile station idle mode cell selection	AR (Argentina)
3	None	1380	None	Communication control technique for a radiotelephone system	VE (Venezuela)
4	None	PCT/NL/95/00055 - 94200236.1	666 550	Data exchange system comprising portable data processing units	EP (European Patent Office)
5	None	006151-4	None	Device and method for generating sync word and transmitting and receiving the sync word in W-CDMA co	None
6	None	00426/DENP/2004	None	Radio resource control-service data unit reception	None
7	None	0006167-0	None	device and method for implementing handoff in mobile communication system with short sync channel	None
8	None	P-003071	335	Frequency synthesizer for broadcast telephone system having multiple assignable frequency channels	None
9	None	3045/86	67260	Subscriber unit for wireless digital telephone system	None
10	None	77/90	132016	A wireless digital communication system	None
11	None	20-2002-0012097	283799	Physical layer processing for wireless communication system using code division multiple access	None
12	None	15/324391	None	Frequency hopping pattern with reindexing RB for D2D scenario	US (United States)
13	None	61/523,113	None	Method and apparatus for handling additional power backoff	US (United States)

Examples 12 and 13 are interesting, as they refer to serial application numbers of patent applications. Example 12 has a formatting as used at the USPTO. The formatting is different from EPODOC formatting used for harmonized ETSI records and used in PATSTAT. With considerable effort, the patent (application) in question be found in PATSTAT, but this depends on whether it was abandoned by the owner or not, and on whether the application was done prior to March 2012 or not,[30] among other things. Example 12 is a provisional patent application in the USPTO. Such an application will not directly lead to a granted patents, yet may or may not later be referred to as a priority document for a later application. Such provisional patent applications, especially if they never served as priority document later in their life, can be very hard to find in PATSTAT.

Patents similar to examples 12 and 13 can be found in great number in the retrieved ETSI database. A non-exhaustive, quick look showed at least 3600 serial numbers of applications (more than 1200) or provisional applications (more than 2400) at the USPTO alone.

We expect that for many of these patents, the patent owner later on, when a publication number was available, submitted a complete disclosure, resulted in a harmonized record. Yet, the 'old' unrecognized

[30] Only from March 2012 on, the USPTO started to publish patent applications. Patent applications that fall before that publishing regime, and who did not see later grants, do have USPTO serial numbers (nn/nnn,nnn) but are not present in PATSTAT.

application was never retrieved. We manually looked at a dozen of cases where we managed to identify the patent, and found for each of them that this expectation was indeed the case

All in all, our conclusion is that:

- One part of the unmatched records concerns patents for which the owner provided incomplete or erroneous information.[31] Such patents cannot be matched by us in an automatic fashion and doing so manually by us requires an effort that clearly falls outside the scope of this project.
- Another part of the unmatched records concerns patents that are in the early application phase, including provisional patent applications. They may not have developed into real applications or granted patents at all. Or they may have developed into a real application or a real granted patent, but that patent might already be in the ETSI database, while the discloser did not remove the old, unharmonized record.

All in all, we feel that there should not be a too great concern over these unmatched patents in the context of our study, where precision (are all the patents we identified actually disclosed patents) presides over recall (have we missed any disclosed parents).[32]

2.5 Summary of dataset of identified basis patents

The content of our final dataset of disclosed ETSI patents, in terms of the various types of patent families, is shown in Table 7. In ETSI's disclosure process, the disclosure of one single patent suffices to disclose all existing and future members of its patent family (as per the ETSI family definition).[33] Yet, declarants may sometimes disclose multiple members of a patent family.[34]

After having matched with PATSTAT, we can also distinguish cases where two or more basis patents have exactly the same set of priorities[35] (in the previous illustrative pictures, these could be patent 3 and 4, or patents 7 and 8). Removing these duplicates brought our dataset down from 34,711 records to 25,072, which is our final dataset.

As can be seen in Table 7, our final set also covers 25,072 DOCDB families. That is logical; since the DOCDB family definition is smaller than the ETSI patent family and by definition covers all items of the latter, we should find exactly one unique DOCDB family for each identified unique ETSI basis patent. Yet, even though the *number* of DOCDB and ETSI families is the same, their *content* is very different (see Section 2.2), and DOCDB families would have been much less appropriate for the FRAND licensing commitment ETSI seeks to obtain.

As we can also see in Table 7, there are fewer INPADOC families in the final set. This is also logical, since the INPADOC family is more inclusive, a single INPADOC family can encompass multiple ETSI family members. These INPADOC families are, again, of very different *content* than the ETSI families.

[31] Note that, in recent years, ETSI has encouraged declarants to provide additional information for such patents to be able to match them. Processing this input by ETSI is a time-consuming proves, but step by step, the share of unmatched records is decreasing.

[32] Of course, in another context this could be different (e.g. a situation where a patent holder asserts a patent and the question is on the table whether that patent was disclosed at ETSI or not).

[33] "The obligations pursuant to Clause 4.1 above are deemed to be fulfilled in respect of all existing and future members of a PATENT FAMILY if ETSI has been informed of a member of this PATENT FAMILY in a timely fashion. Information on other members of this PATENT FAMILY, if any, may be voluntarily provided.". Clause 4.3 of the ETSI Intellectual Property Rights Policy, in Annex 6 of ETSI (2018a).

[34] *"Other patents of the patent family may have been explicitly disclosed on the declaration by the declaring company."* ETSI (2011), Section 5.3.3.

[35] This is the case of they belong to the same DOCDB family.

Table 7. Characteristics of final dataset of ETSI disclosed patents

Type	Unique occurrences in final dataset	Note
ETSI disclosed ('basis') patents, each with their unique ETSI patent family	25,072	This are ego-families, valid from the perspective of the disclosed ('basis') patent. Families may overlap in terms of family members they cover.
DOCDB patent families	25,072	These are mutually exclusive families (so can never overlap)
INPADOC families	23,643	These are mutually exclusive families (so can never overlap)

2.6 Creating patent families for our analyses

As discussed above, our analysis will use the ETSI patent family definition where appropriate. We created such families by running SQL queries in the PATSTAT database, using the ETSI basis patents (disclosed patents) as the starting point. We implemented that in the following way:

Patent A belongs to the patent family of the focal patent when at least one of the following conditions is satisfied:[36]

1. Patent A is listed as a priority by that focal patent,
2. The focal patent is listed as a priority by patent A,
3. Patent A and the focal patent share one or more priorities.

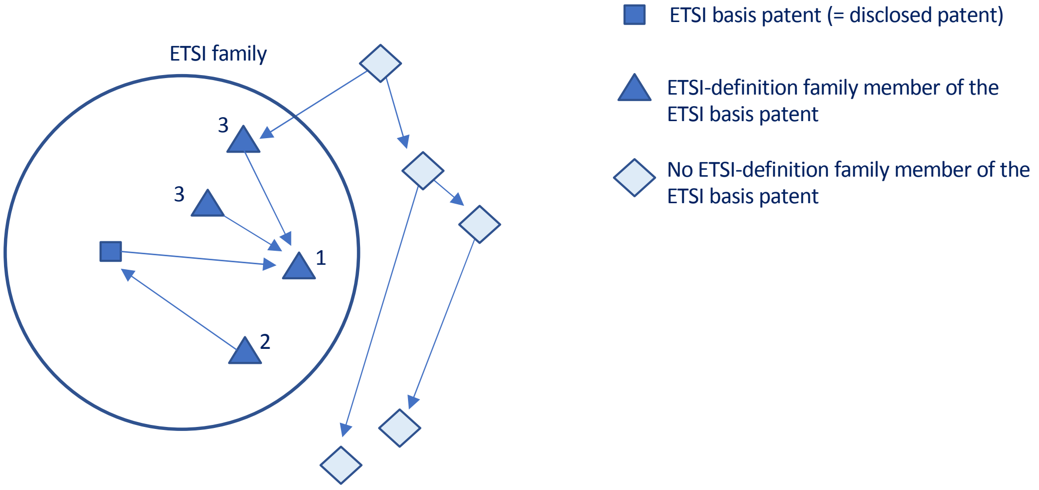
In order to double-check our mechanism, we manually inspected a number of patent families this way, and found that our algorithm creates the same patent families as those that are shown when the specific disclosure (ISDL) and disclosure number manually are looked in the 'declaration search' took of the ETSI IPR online database.[37]

In Figure 6, a number of patents and their priority relations are shown. The square patent is the ETSI base patent, and the triangle patents are ETSI definition family members of that ETSI base patents, where the numbers refer to the above condition that makes them family member. The diamond patents have a direct or indirect priority relation to ETSI definition family members of that ETSI basis patents but are not ETSI definition family members of that ETSI basis patents itself.

[36] When carrying out or work, we came across some cases in which a patent was in the same DOCDB family as the focal patent, yet none of the three conditions was met. We believe this may be the case of some missing or inconsistent priority data in PATSTAT. To ensure capture the full ETSI family, we add all patents that are in the same DOCDB family of the focal patent as well.

[37] We do note that in the ETSI retrieved database itself there are sometimes different lists of priorities listed for records for the same patent (see Table 4). We also note that PATSTAT is constantly updated, and in a newer version, some additional patents may show up in the family construction (newer patents that list one of the priorities we search for). In that sense, our results will not necessarily be *identical* to the families shown via the ETSI IPR 'declaration search' tool (which we believe is created once, and not recreated every time users consults the database).

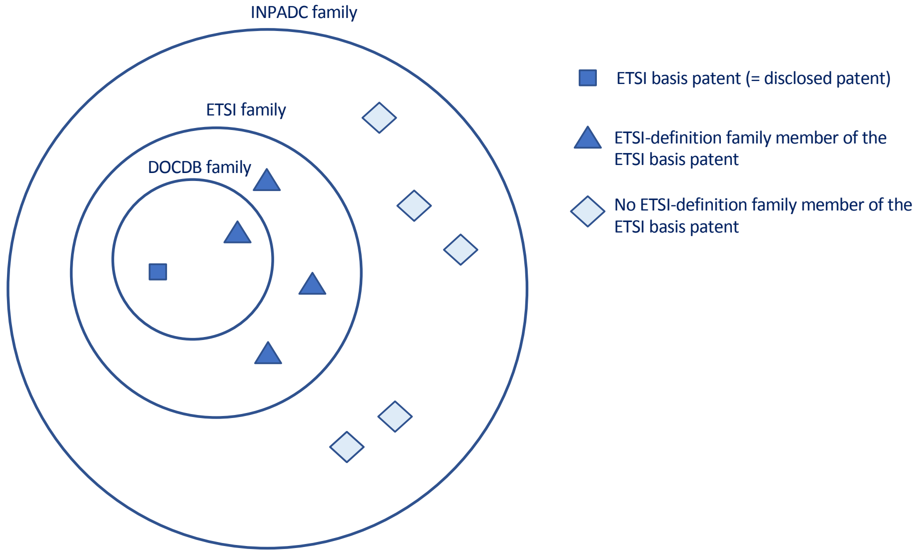
Figure 6. Example of an ETSI patent family



Note: Numbers shown on the basis of which of the three conditions they became family member (see above)

The same patents are also shown in Figure 7, but now also showing which patents are in the same INPADOC family and in the same INPADOC family.

Figure 7. ETSI patent family compared with DOCDB and INPADOC patent family



2.7 Distinguishing standards in the ETSI disclosure database

There is a very understandable desire to distinguish patents disclosed at ETSI for what one may call different standards (e.g., 2G, 3G, 4G and 5G). And many reports are making such a distinction, or at least claim to do so.

Yet, this is a task that is not easily accomplished, for several reasons:

1. The 'ETSI project' field in the ETSI disclosure process (and hence also in the database) is of free format, and a discloser can enter its own, new descriptions and acronyms. In our retrieved ETSI dataset, there are 333 different ETSI Project entries, with no apparent structure.[38]
2. The selective ETSI Projects that do cover the bulk of the ETSI disclosures are very broad. They refer to activities in 3GPP, and such activities may cover 2G, 3G, 4G and 5G at the same time.
3. ETSI disclosures may contain information in the field called 'standards' (which is also known as the 'illustrative part of the standard' in ETSI parlance. But this field is not mandatory for a discloser to fill in, and as of now, less than 25% of all disclosures in the ETSI database provide this information. And if they do so, it is not easy to translate that into technology generations either. The information in the field is unharmonized, and usually refers to 3GPP documents (Technical Specifications) than cannot 1:1 be linked to 3G, 4G and 5G either.[39]
4. Generally, ETSI disclosures specify a specific Technical Specifications (TS) document of 3GPP for which the patent is potentially essential. A TS may (and most often does) relate to multiple generations (2G, 3G, 4G or 5G). Recently, the ETSI secretariat produced a spreadsheet that links specific TS documents, including their version number, to these four generations (ETSI, 2019). This spreadsheet shows, for instance, that TS24.011 V7.1.0 relates to 3G and 4G, whereas TS24.011 V8.0.0 relates to 3G, 4G but also 5G. Most ETSI declarations include a TS number, and some (but certainly not all) also include a version number. Suppose we have a disclosed patent that includes both: TS24.011 V8.0.0. In that case, we can assume the patent is not believed to be potentially essential to 2G, but still we cannot infer anything whether it is believed to be potentially essential to 3G, to 4G, to 5G, or any combination of these.
5. What is referred to as 2G, 3G, 4G and 5G relates especially to radio interfaces, but the ETSI/3GPP standards contain much more than radio technologies, such as technologies in routing (for which there are also several generations, but not 1:1 linked to radio technologies). They may relate to security, or the (U)SIM card. Patents may be believed to be technologies other than radio interfaces, for instance and then cannot be easily linked to 2G, 3G, 4G and 5G.
6. Developments on different generations (e.g. 4G and 5G) to some degree take place in parallel. So, it is not possible to link to different generations merely on the basis of the moment of the invention (or patent application), or the moment of disclosure.

All in all, we believe that, given the available information at the time we performed our analyses, it was impossible to break down ETSI disclosed patents for different generations (2G, 3G, 4G and 5G) in a reliable way.

[38] When making selections in the Dynamic Reporting tool at ipr.etsi.org, a hierarchical tree is available for search purposes, but this tree contains much more entries that are "unharmonized' than not, and the actual output in the retrieved data does not have that hierarchical structure.

[39] Furthermore, 3GPP, in its specification series, only distinguishes between three broad series: (a) 3G and beyond / GSM (R99 and later), (b) GSM only (Rel-4 and later) and (c) GSM only (before Rel-4).

3 A landscape analysis of ETSI disclosed patents

Section 2 carefully described the procedure we followed to create a list of all the unique patents disclosed to ETSI and to identify all the members of the family of the disclosed patents. As discussed above, we then matched the disclosed (basis) patents to PATSTAT to gather additional and up-to-date data about these patents. In particular, we are interested in obtaining information about the timing of patent filing and disclosure, on the disclosing entities, information about the technological and geographical distribution, and proxies for the quality of the disclosed patents. This section describes the construction of the variables capturing this information and then presents descriptive statistics allowing us to have a first look at the disclosed SEP landscape.

3.1 Construction of variables of interest using the PATSTAT database

Several dates are important in the lifetime of a patent: the date of filling at the patent office, the date of publication of the application and search report, the date of grant, etc. The priority date is the date of the first filing of the application in any office, and in terms of dates available in patent databases, it is the date that comes the closest to the actual time of invention. Prior art can only be considered in the examination process if it is made public before the priority date.[40] The date when the patent was disclosed at ETSI as potentially standard-essential is the date of disclosure. The disclosure lag is the difference between these two dates: how long after the priority date was the patent disclosed at ETSI.

The number of applicants indicates a collaboration between organizations, while the number of inventors in a patent is related to the size of the research team. Both signal the size of the research project, which is related to the probability of developing “big innovations” (Mariani, 2004).

The number of distinct IPC classes (at four digits) reported in the patent document captures the technological breadth of a patent, that Lerner (1994) shows is related to its value.

The family size is the number of patents in the same family, while the number of legislations is the number of different patent offices of the applications in a family. This indicator captures the market scope of an invention (Lanjouw et al., 1998; Harhoff et al., 2003). As multiple applications are costly, applicants will be willing to bear these additional costs for protection only for valuable innovations. The family size and number of legislations can be calculated for any family definition (e.g., DOCDB, INPADOC, or the ETSI family definitions).

To establish its prior-art, patent documents refer to previous patents and scientific literature. The number of references to earlier patents measures the degree of novelty of an invention and, therefore, it is associated with patent quality (Criscuolo and Verspagen, 2008). Self-citations are a particular class of backward citations as they are references to patents of the same applicant. These citations identify within-firm cumulative innovations, which might indicate particularly valuable lines of innovation. A patent application can have several publications (identified with distinct kind codes), so the number of backward citations is the number of unique citations to patent documents from any of those publications.

The number of references to non-patent literature (NPL), that are mostly scientific literature, are generally used to measure the link between science and technological development (Narin et al., 1997). However, patents related to basic research are more valuable because covering more complex and fundamental knowledge (Cassiman et al.; 2008; Branstetter, 2005). The number of NPL references is the number of unique non-patent documents cited in any publication of a given application.

Since the seminal work by Trajtenberg (1990), the number of citations (‘forward citations’) received by a patent has been extensively used as a measure of technical quality (Hall et al., 2005; Harhoff et al., 2003). The logic is quite simple and similar to the one of using citations to evaluate the quality of scientific publications. Important patents will foster follow-up innovations that will cite the originating patent. The number of forward citations is the number of unique patent publications that cite any publication associated with a given application.

40 The priority date can be found in PATSTAT under the “earliest filling date” column of tls201_appln table.

The next variable is the number of claims reported in the patent. Claims define the boundaries of the exclusive right and determine what can be legally protected and enforced. Claims have been used as a measure of patent value (Lanjouw and Schankerman, 2001, 2004) because they relate to the breadth of an invention and, therefore, the possibility of broader application. Furthermore, as in several patent systems fees increase with the number of claims, they are directly linked to costs that an applicant is incurring. When a patent application has several patent publications associated, we take the highest number of claims reported in a publication of this patent.

Finally, patents are only enforceable if they are granted. The value of a patent that has been granted is higher than that of a patent application (Sherry and Teece, 2004). Our last variable is a dummy that indicates whether the patent has been granted.[41]

3.2 A first look at the data

As discussed in Chapter 2, we identified 25,072 disclosed patent families (following the ETSI family definition). In this section, we provide descriptive statistics on the basis patents associated with these families (i.e., the patent provided by the disclosing firm), using the bibliographic information as recovered through PATSTAT (2018b) of the patents that were actually disclosed in the disclosure letters to ETSI by the patent owners. If one of the patent families (following the ETSI family definition) we identified has more than one member disclosed at ETSI, we recover the patent-level information of the patent with the earliest disclosure date.[42]

Table 8 reports general summary statistics about our main variables of interest. Following sections discuss a few of these dimensions in detail.

3.3 Timing of disclosure, filing and age of the disclosed patent

The first important dimension to analyse relates to the timing of disclosure of the SEPs to ETSI. As Table 8 shows, about 76 percent of the patents disclosed to ETSI, were first disclosed after 2010, whereas less than 25 percent of the disclosed patents were disclosed before 2010. Figure 8 further explores the distribution of disclosed SEPs over time, showing the distribution over the year of the disclosure.[43] It shows that the distribution is particularly skewed towards the end of the period taken into account, with more than nine thousand SEPs disclosed in 2017 and 2018 alone.

The distribution of disclosed SEPs over time is somewhat less skewed if we consider the year of first priority of the patent (Table 8). In this case, the majority of the disclosed SEPs have a priority date that goes back to before the year 2010, whereas about 40 percent of the disclosed SEP families we consider was first filed from year 2010 onwards. Nevertheless, Figure 9, which displays the distribution in more details, shows that also in this case the distribution is quite skewed towards the second part of the time period considered, with more than 17 thousands disclosed patents and patent applications having a priority date from the year 2007 onwards, and almost 2,400 in the year 2016 alone.

[41] This information comes directly from the column “granted” in table tls201_appln.

[42] Whenever we speak of ‘number of patents’ or ‘number of SEPs’ in this section, we mean the number of patents disclosed to ETSI, but we count only one patent for patent families (following the ETSI family definition) that have more than one patent disclosed. The patent-level information in the latter case, comes from the earliest disclosed patent.

[43] For this figure, 19 patents were not included as their records gave ‘January 1, 1900’ as a disclosure data and have to be considered as data errors. Given the size of our data set, discarding 19 patents does not significantly impact results.

Table 8. Summary statistics (data set: the 25,072 disclosed basis patents, coming from different jurisdictions)

Variables	Mean	SD	Median	Min	Max
Nb. of inventors	3.12	1.8	3.0	1	19
Nb. of applicants	2.07	1.9	1.0	1	20
Age of the disclosed patent	5.26	3.9	4.0	0	34
Nb. of IPC_4_DIGIT	2.23	1.6	2.0	1	30
Nb. of legislations (DOCDB)	5.74	4.3	5.0	1	29
Nb. of legislations (ETSI family definition)	5.72	4.6	5.0	1	29
Nb. of legislations (INPADOC)	6.61	5.1	5.0	1	33
DOCDB family size	8.14	10.6	6.0	1	269
ETSI family size	10.4	20.3	7.0	1	640
INPADOC family size	30.24	190.5	7.0	1	2938 (*)
Nb. of claims	21.68	15.0	19.0	1	264 (*)
Nb. of backward citations	10.30	22.8	4.0	0	1176 (*)
Nb. of non-patent literature	3.67	9.7	1.0	0	128
Nb. of forward citations	24.61	59.6	4.0	0	2478 (*)
Granted	0.73		1.0	0	1
Disclosure year	**Mean**	**SD**	**Median**	**Min**	**Max**
2000-2004	0.08		0	0	1
2005-2009	0.15		0	0	1
2010-2014	0.32		0	0	1
Post-2015	0.44		0	0	1
Priority year	**Mean**	**SD**	**Median**	**Min**	**Max**
Pre-2000	0.13		0	0	1
2000-2004	0.14		0	0	1
2005-2009	0.31		0	0	1
2010-2014	0.27		0	0	1
Post-2015	0.14		0	0	1
N=25,072					

Note: *(*) these are outliers in the data set, but not data errors; this table actually reports on the data as present in the PATSTAT database. In fact, some recent research papers investigate patents with extremely high numbers of backwards citations (Kuhn et al., forthcoming).*

Figure 8. Number of patent families by year of first disclosure (data set: the 25,072 disclosed patent families).

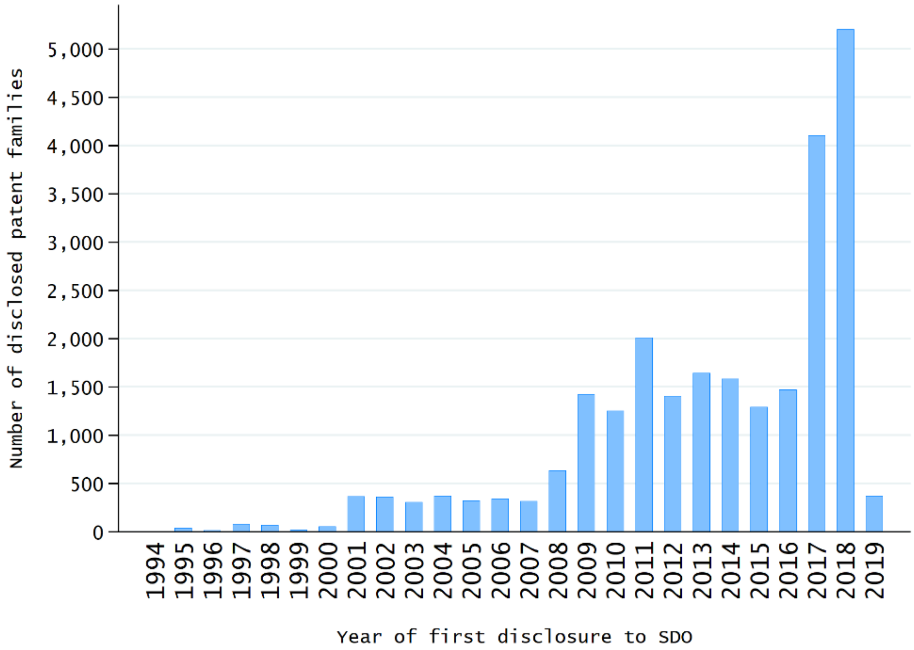


Figure 9. Number of disclosed patent families by earlies patent priority date within that family (data set: the 25,072 disclosed patent families).

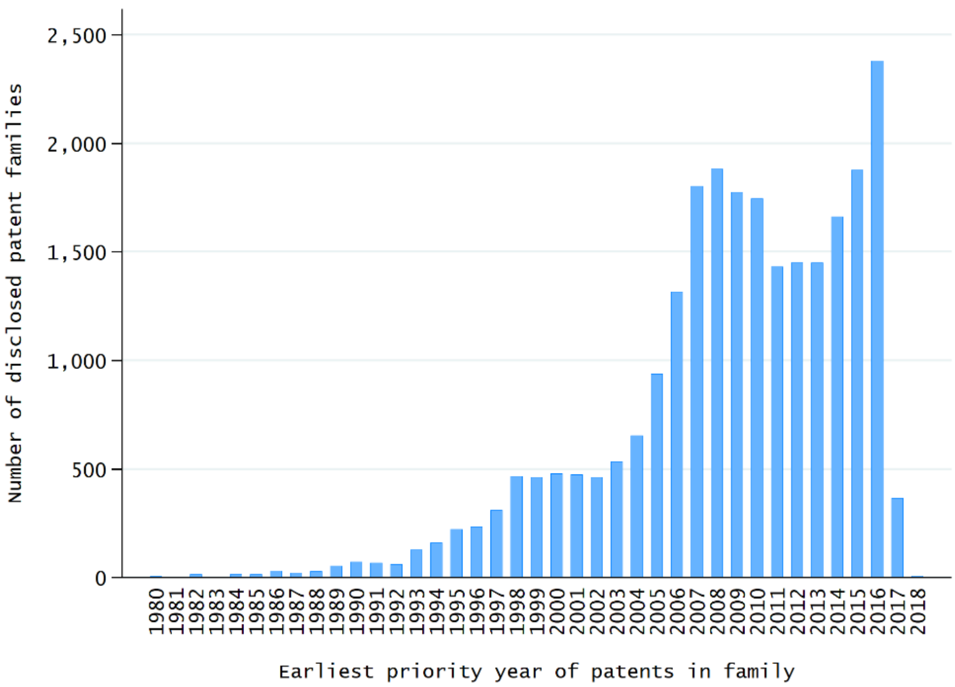
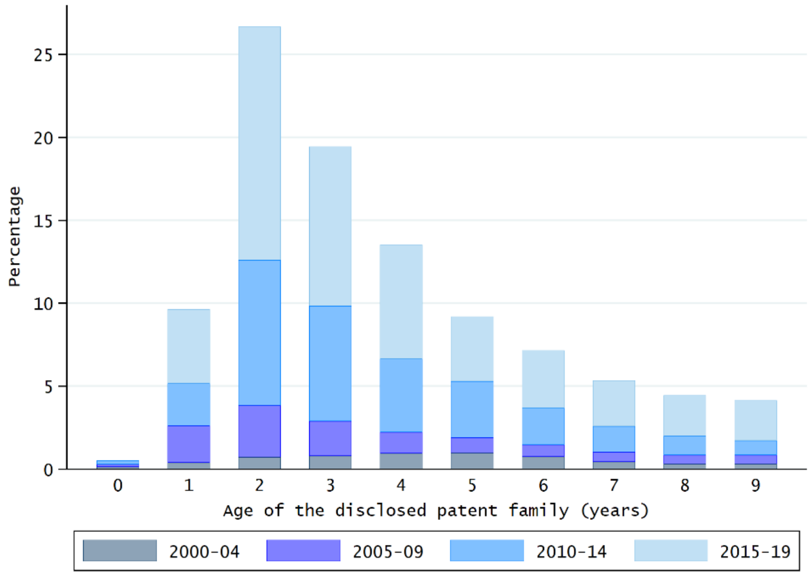


Table 8 **(**on page 23) also showed the *difference* between the year of first priority and the year of the first disclosure, which we call the “Age of the disclosed patent”. The average is about 5.2.[44] Also this distribution is skewed: the median age of the disclosed patent is four years, the longest age we encounter in the data is 34 years (note that this patent could not be enforced any more at the time it was disclosed). 90 percent of the disclosed SEPs has an age that is younger than 10 years. Figure 10 shows detail of the age distribution for disclosed patents (within the 90 percentile of age), where the age is the difference the two variables we just discussed: the year of first disclosure and the earliest priority year of the patents in the family. The figure also reports the distribution over different time windows of the year of first disclosure. The mode of the distribution is 2 years.

Figure 10. Share of disclosed patents by age of the disclosed patent, in years (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).



[44] 11 patents had negative disclosures lag which we believe is due to an error in the data and were not considered.

3.4 Technological trends

As discussed in Section 3.3, the variable *Nb of IPC_4_DIGIT* reports the number of distinct IPC subclasses (i.e., four-digit level) assigned to disclosed SEPs. Table 8 **(**on page 23) already reported that, on average, disclosed SEPs are assigned to 2.2 different IPC subclasses. Figure 11 shows how almost 80 percent of the IPC four-digit classes assigned to disclosed SEPs belongs to only three subclasses: *H04W*, *H04L*, and *H04B*. Not surprisingly, these subclasses all belong to the *H04* IPC class that covers *Electric Communication Technique.*

Figure 11. Share of disclosed patents by main IPC subclasses
(data set: the 25,072 disclosed basis patents, coming from different jurisdictions).

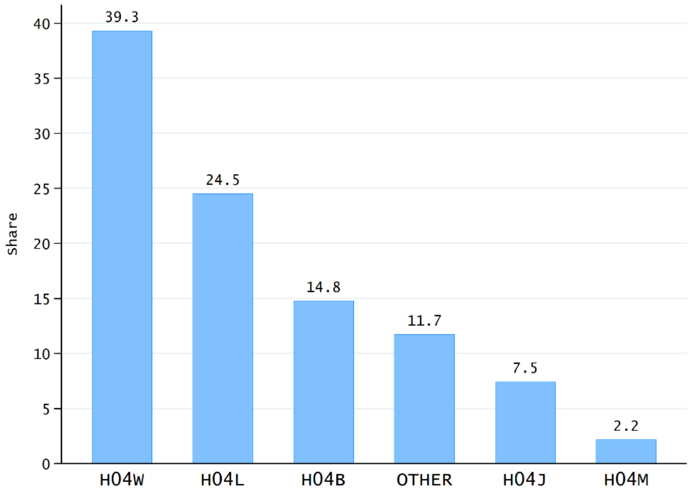


Figure 12 illustrates the evolution of the distribution of the three most frequent subclasses plus a residual subclass, over the year of disclosure. As the figure shows, the concentration in the *H04W* subclass becomes especially evident with the patents disclosed after 2008.

Figure 12. Density distribution of the main IPC subclasses by year of disclosure
(data set: the 25,072 disclosed basis patents, coming from different jurisdictions).

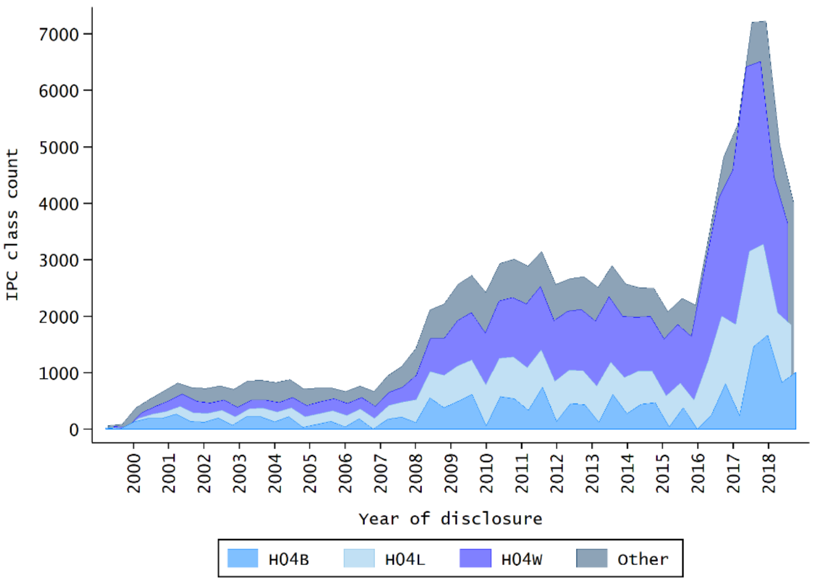
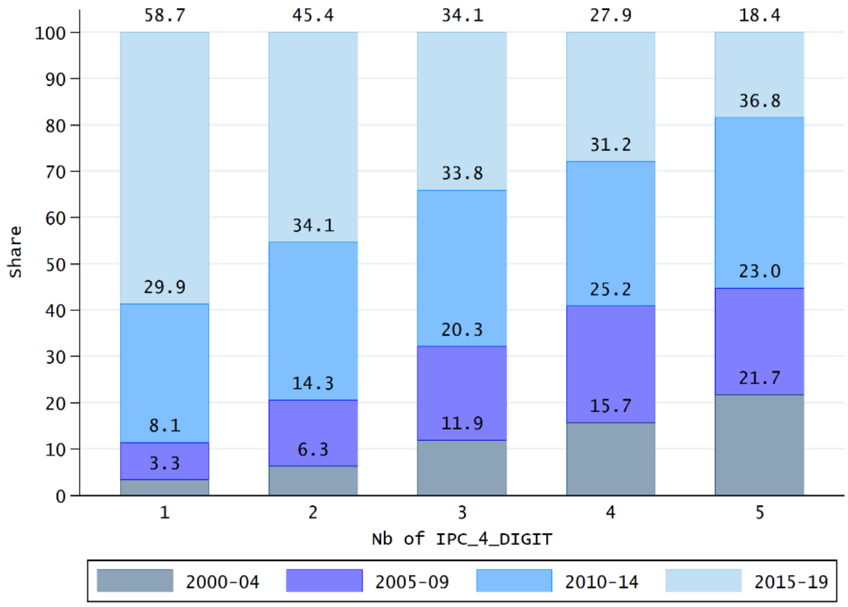


Figure 13 shows that patents disclosed in more recent years are less diversified over IPC subclasses. Patents disclosed since 2010 represent almost 90 percent of the patents assigned to a single IPC subclass, but less than 60 percent of the patents assigned to more than four distinct IPC subclasses.

More specifically the most frequent IPC subclass observed in the data is H04W, which covers Wireless Communication Networks, whereas H04B and H04L covers Transmission and Transmission of Digital Information respectively. Overall, less than 8 percent of the subclasses assigned to disclosed patents do not belong to the H04 class, with the most frequent classes outside H04 being G06 (Computing, calculating, Counting) and H03 (Basic Electronic Circuitry)

Figure 13. Number of different IPC subclasses associated with our patents (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).

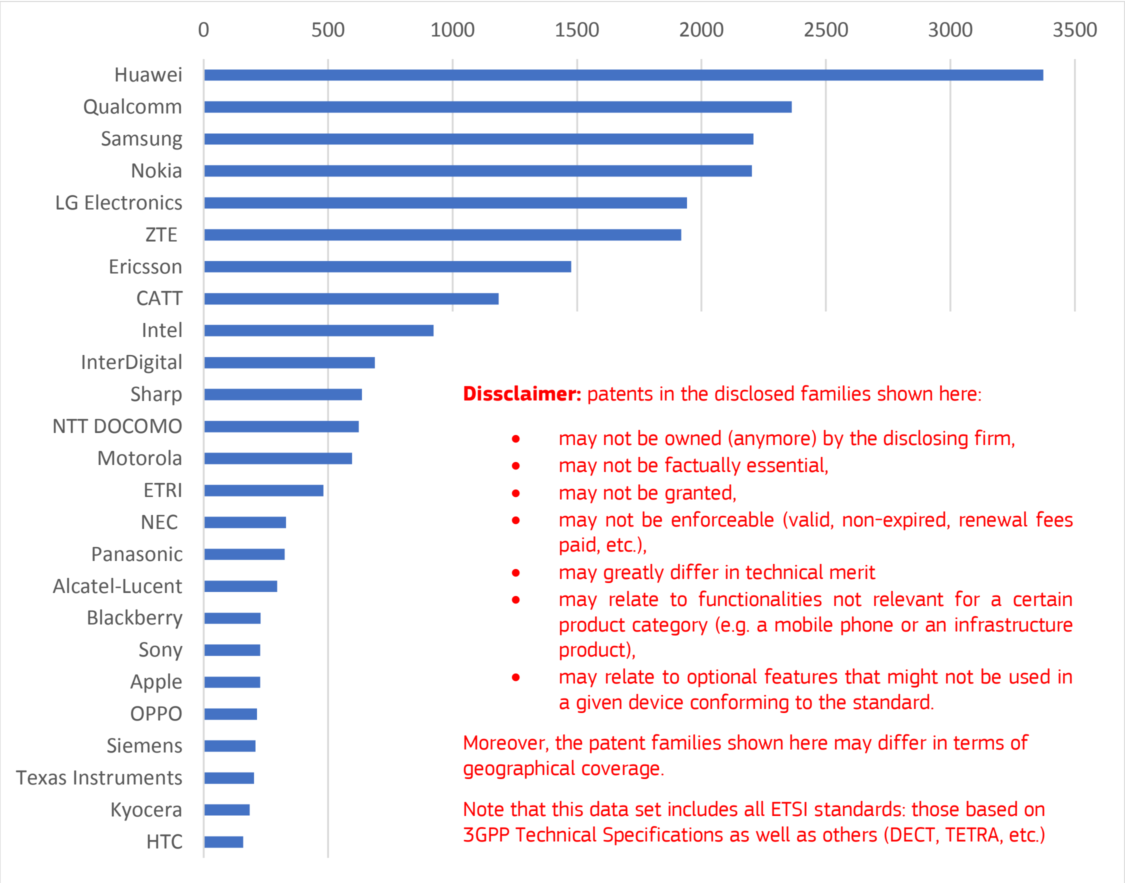


3.5 Disclosing firms

The ETSI IPR database contains the name of the "declaring company",[45] that is, the firm disclosing the IPR in question. This data reflects the ownership *at the time the IPR was disclosed to ETSI*.[46] We analysed this data for the 25,072 patent families (following the ETSI family definition) in our dataset, where we looked up the *earliest declaring firm* for the family in question.[47] As some companies used different names when declaring their patents, we performed name cleaning, but we did not consider mergers or acquisitions (M&A's), transfer patent portfolios or parts thereof, ownership changes resulting from bankruptcy, and so on, as that goes beyond the scope of this study. Details on our name cleaning are provided in Appendix 1. Figure 14 shows our findings for the 25 largest companies. More detailed data, for all 67 companies that have disclosed 10 or more patents, is given in Appendix 2.

It is important to keep in mind, however, that this data does not imply *current* ownership. Also, it does not tell whether and in what part of the world patents in this family are enforceable, and what the value of those patents is. Finally, as for all the data in this chapter, it does not tell us whether the patent family is factually essential.

Figure 14. Number of disclosed patent families (following the ETSI family definition) by earliest disclosing company; shown for the 35 largest companies; full time period (1990-2019). (data set: the 25,072 disclosed basis patents, coming from different jurisdictions)



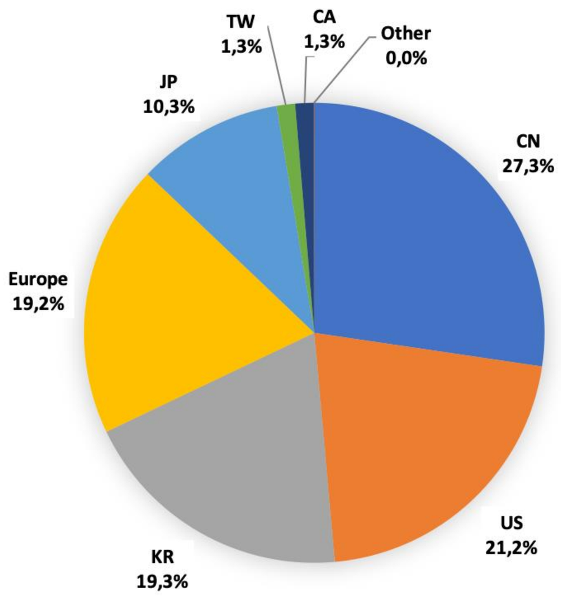
[45] Even though ETSI uses the word 'company' (and we follow that convention in our report), note that this may be any type of entity (e.g., public research organisations, associations). For the term 'declaring', see also Section 1.2

[46] While technically the ETSI IPR policy does allow firms to inform ETSI of IPR held by third parties, this never seems to be the case for the records in the database. In the on-line ISDL declarations, we invariably see stated that the "The declarant and/or its affiliates are the proprietor of the IPR(s)."

[47] It seems there might be instances where after a disclosure was made by a 'new' owner of a patent, the record of the original patent owner was longer no present in the 'dynamic reporting' tool of the ETSI IPR database. We discussed this with ETSI staff and the exact cause of this is not known, neither the extent at which this may occur.

We also looked at spatial patterns in patent family disclosure. In order to do so, we determined the 'home country'[48] for all 148 (cleaned) declaring company names in the database. As can be seen in Figure 15, the largest number of patent families (following the ETSI family definition) are disclosed by Chinese firms (27.3%), followed by US firms (21.5%), South Korean firms (19.3%) and by European firms (19.2%).

Figure 15. Share of number of disclosed patent families (following the ETSI family definition) by home country/region of disclosing firm (data set: the 25,072 disclosed basis patents, coming from different jurisdictions)



Notes: Europe includes companies from EU countries, Switzerland and Norway.
The 'other' category includes firms from Israel, Turkey, and New Zealand.

[48] We consider the 'home country' as the principle country of the firm, and typically it is where their factual headquarters are (which is not necessarily where the company is legally registered). In case of doubt, we consulted the ISDL declarations the company made to ETSI (including the underlying facsimile documents). For the vast majority of entries, the home country is relatively easy to establish. However, there are some exceptions, typically smaller companies. An example is Vimatix: on the facsimile underlying the ISDL statement submitted to ETSI, this company is called 'Vimatix (Technologies) Ltd', the provided C/O postal address is in the US, yet the fax is sent from Israel, and the stamp of the firm at the signature page reads "Virgin Islands'.

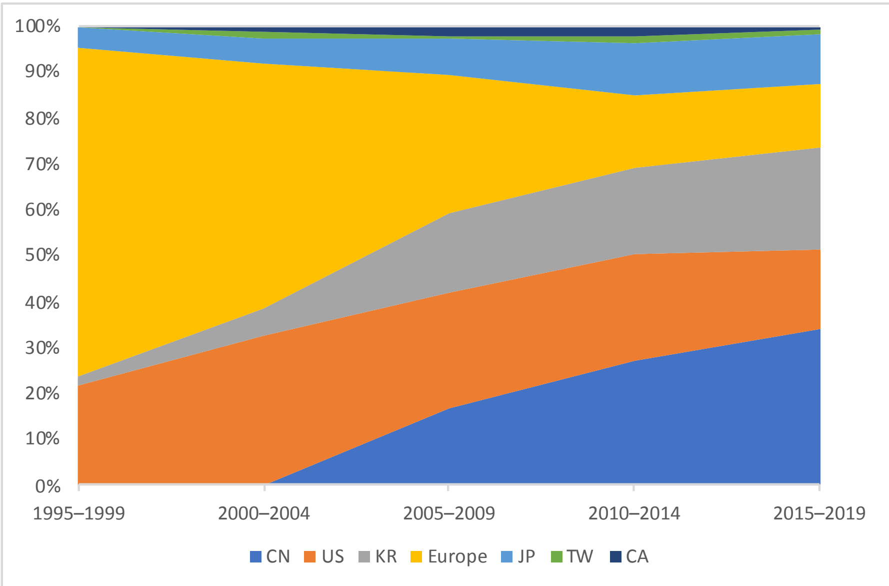
These spatial patterns also change over time (Table 9, same data visualized in Figure 16). Here we see that it was initially the European companies that were leading in the number of disclosed patent families (following the ETSI family definition).[49] Other countries/regions in the world increasingly caught up, and Chinese companies took over the lead from 2010 on.

Table 9. Number of disclosed patent families (following the ETSI family definition) by home country/region of disclosing firm, over time (data set: the 25,072 disclosed basis patents, coming from different jurisdictions)

Country/ region	**1990–1994**	**1995–1999**	**2000–2004**	**2005–2009**	**2010–2014**	**2015–2019**	**Full period**
CN	0 (0%)	0 (0%)	0 (0%)	506 (17%)	2135 (27%)	4214 (34%)	6855 (27%)
US	6 (55%)	48 (22%)	474 (32%)	767 (25%)	1835 (23%)	2194 (18%)	5324 (21%)
KR	3 (27%)	4 (2%)	81 (6%)	520 (17%)	1495 (19%)	2747 (22%)	4850 (19%)
Europe	2 (18%)	158 (71%)	778 (53%)	921 (30%)	1259 (16%)	1698 (14%)	4816 (19%)
JP	0 (0%)	9 (4%)	78 (5%)	245 (8%)	891 (11%)	1347 (11%)	2570 (10%)
TW	0 (0%)	0 (0%)	24 (2%)	13 (0%)	124 (2%)	163 (1%)	324 (1%)
CA	0 (0%)	1 (0%)	16 (1%)	62 (2%)	164 (2%)	79 (1%)	322 (1%)
Other	0 (0%)	2 (1%)	8 (1%)	0 (0%)	0 (0%)	1 (0%)	11 (0%)
TOTAL	11	222	1,459	3,034	7,903	12,443	25,072

Notes: Europe includes companies from EU countries, Switzerland and Norway.
The 'Other' category includes firms from Israel, Turkey, and New Zealand.

Figure 16. Share of disclosed patent families (following the ETSI family definition) by home country/region of disclosing firm, over time (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).



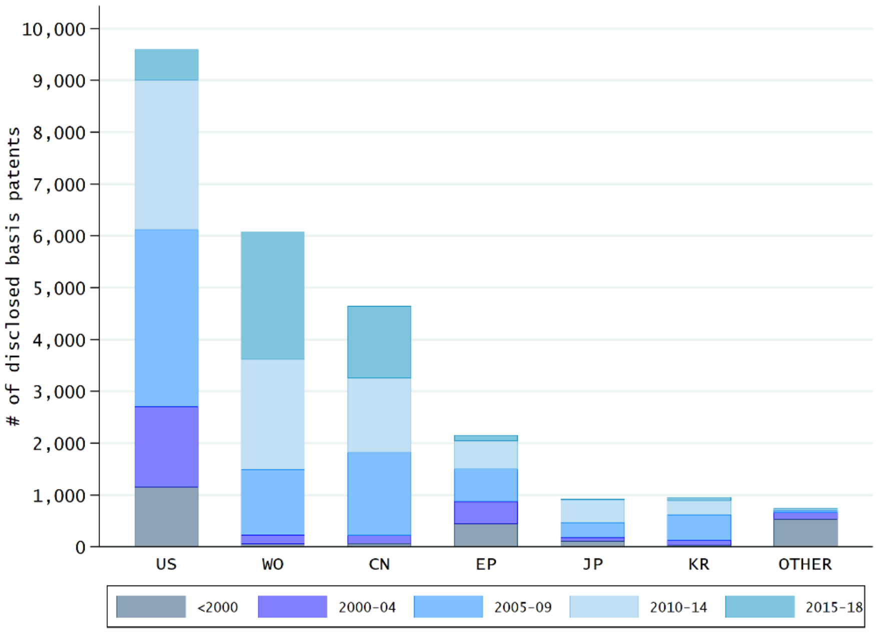
Note: the 1990-1996 time frame is omitted in thus picture, as there are few observations in that time slot.

[49] Even though companies from the US were the largest disclosers of patents in the period from 1990 to 1994, there are only 11 disclosed patent families in this period.

3.6 Patent authorities

Another interesting dimension concerns the patent authorities of the basis patents which are disclosed to ETSI. Figure 17 reports on their distribution. Most of the disclosed basis patents (38%) were first filed at the USPTO (US). A substantial amount of applications disclosed as SEPs were directly filed as international (PCT) applications and then report WIPO (WO) as the filing authority. More than 4,500 of the disclosed patents were first filed at the Chinese patent authority (CN), and less than a 1,000 at the Korean and Japanese patent authorities. Figure 17 also allows us to gather some insights about how this distribution evolved over time. In particular, the figure clearly shows that disclosed basis patents filed at the Chinese patent authority come overwhelmingly from more recent years, with only a minority of disclosed SEPs filed in China before the year 2000. The distribution over time of international applications follows a similar trend.

Figure 17. Distribution of disclosed patents by patent authority and time of disclosure (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).



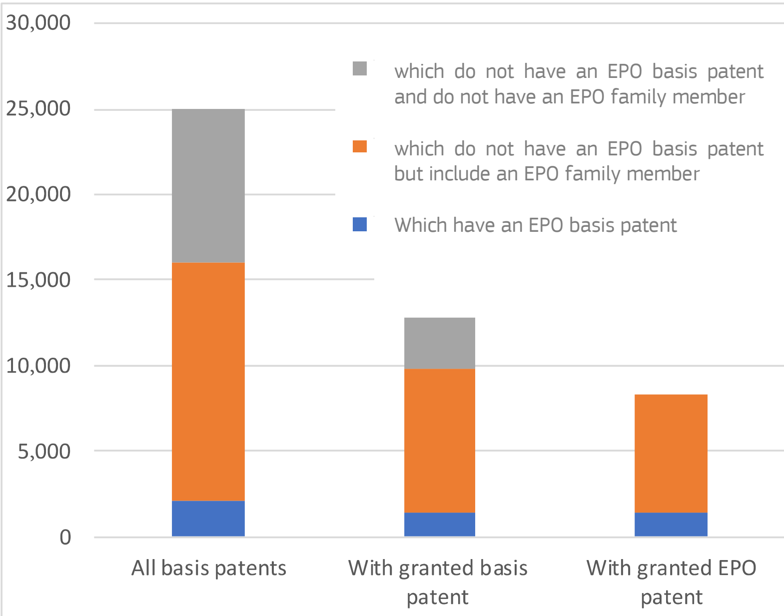
Remarkably, only 2,151 basis patents (8.6 percent of the full sample) were first filed at the European Patent Office (EP) as can be seen in Table 10, Column 2. Nevertheless, 13,938 basis patents filed at authorities different from the EPO (54 percent of the total) were part of an ETSI family that included at least one EPO patent, whereas 46 percent of the disclosed basis patents were not related to any EPO patent. Table 10, Column 3, reports similar figures for the sample that exclusively consider granted disclosed basis patents. Table 10, Column 4, finally, shows the number of families that include a granted EPO family member.[50] Figure 18 visualizes the data in Table 10.

[50] Note that (i) ETSI families are non-exclusive, so the same EPO patent can belong to more than one family (so the final number to be investigated in an assessment system could be a bit lower), (ii) ETSI families may include more than one EPO patent (so it can also be somewhat higher).

Table 10. Number of disclosed patent families (following the ETSI family definition): total, with granted basis patent and with granted EPO patent.

Patent families	Total number	With granted basis patent	With granted EPO patent
Total number of disclosed families (patent families following the ETSI family definition)	25,072 (100%)	12,760 (100%)	8,320 (100%)
.. which have an EPO basis patent	2,151 (8.6%)	1,449 (11.4%)	1,449 (17.4%)
.. which do not have an EPO basis patent but include an EPO family member	13,938 (55.6%)	8,369 (65.6%)	6,871 (82.6%)
.. which do not have an EPO basis patent and do not have an EPO family member	8,983 (35.8%)	2,942 (23.1%)	n/a

Figure 18. Number of disclosed patent families (following the ETSI family definition): total, with granted basis patent and with granted EPO patent.



Finally, we note that the above numbers are for the entire ETSI IPR database. The lion's share of these patent families is about ETSI/3GPP cellular standards 2G (GSM), 3G (UMTS/W-CDMA), 4G (LTE) and 5G. The total number of families declared for other ETSI standards (such as TETRA, ERMES, and the various DVB standards), is very low: we counted appropriately 380 families[51] out of our total of 25,072 families, and another 610 families for which the database field "standard" was equal to "none" and which may, or may not, relate to cellular standards.

[51] Because the information in this field of the database is not harmonized and not always unambiguous, it is difficult to come to definitive numbers.

3.7 Patent family size, claims, and backward and forward citations

Figure 17, in the previous section, only reported on the authority where the patent disclosed in the ETSI disclosure was first filed. Clearly, the patent applicant then may have sought patent protection in other jurisdictions, by extending the application to other patent authorities. As discussed in Section 3.1, the number of jurisdictions in which an invention is protected is correlated with the economic value of the invention (Lanjouw et al., 1998; Harhoff et al., 2003), as filing patents in multiple jurisdictions is costly. Table 8 **(**on page 23) reported the average number of legislations in which the inventions disclosed in the disclosed patents are protected. As discussed above, we consider two types of families to count the number of jurisdictions: *DOCDB* and *INPADOC*. As the table shows, the average DOCDB family of the disclosed patents spans 5.7 jurisdictions, as does the ETSI family, whereas the average INPADOC family spans 6.6 jurisdictions. Figure 19 reports the distribution of disclosed SEPs by the number of legislations for all three kinds of patent families within the 98th percentile. The distributions are overall quite similar, although, as expected, the INPADOC-based count appears to have a longer tail.

Figure 19. Distribution of disclosed patents by number of legislations, for the three different patent family definitions (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).

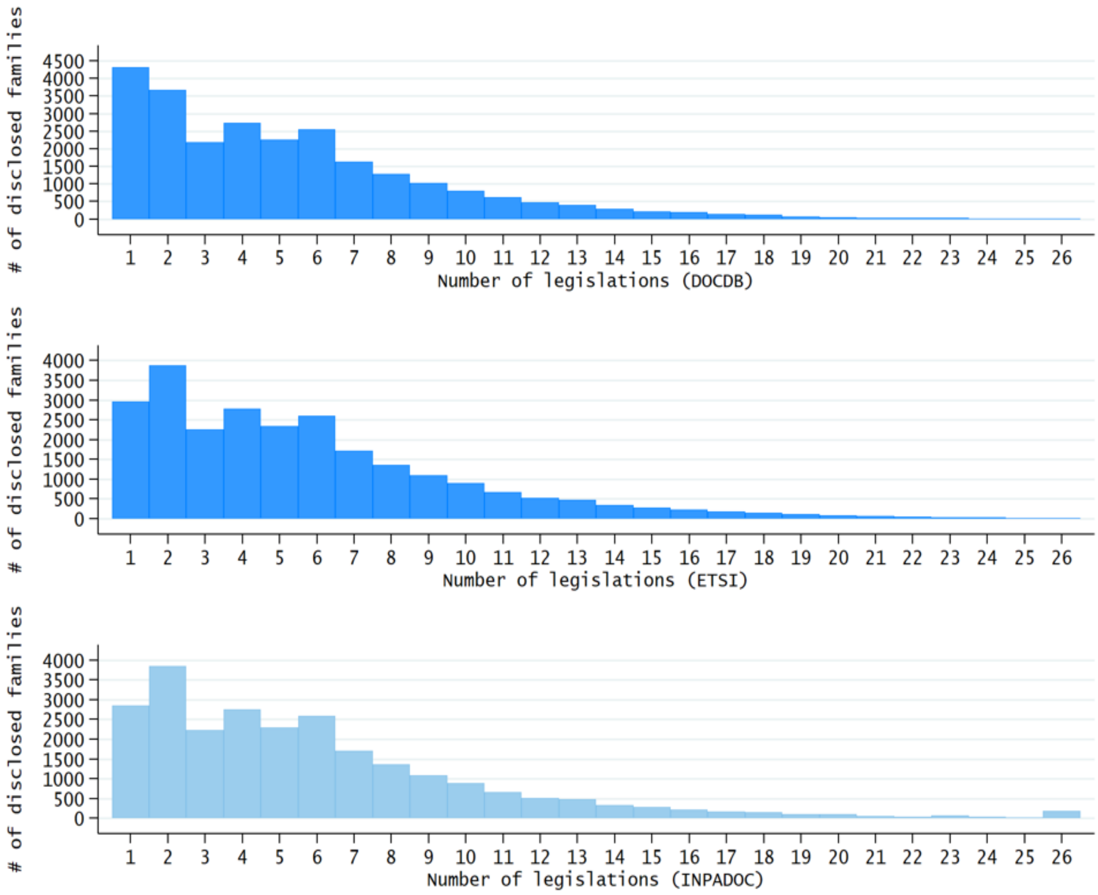
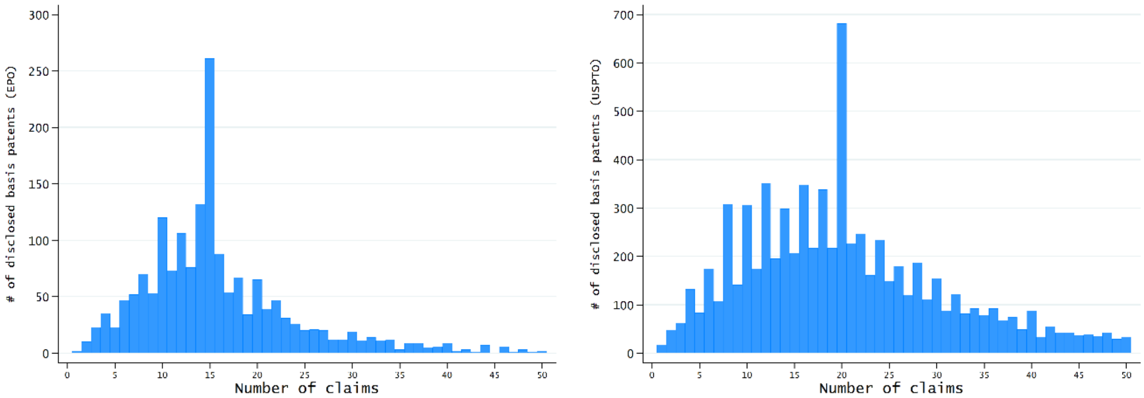


Table 8 **(**on page 23) also reports information about the average size of the DOCDB patent family, ETSI patent family and the INPADOC family of the disclosed patents. Given the substantial differences in the definition, we observe an important difference in the average of these two variables, with the average DOCDB family having 8.1 members, the average ETSI family having 10.4 members, and the INPADOC more than 30 members.

Another variable that has been used in the patent literature to partially capture the value of patent relates to the number of claims included in the patent. For measurement, we use the maximum number of claims associated with one of the patent publications documents. Table 8 (on page 23) already reported that, on average, patents disclosed as SEPs to ETSI have about 21 claims.[52] However, the number of claims included in a given patent application is heavily affected by the institutional environment of the patent authority where the patent application was filed. To draw a more complete picture of the distribution of the number of claims, Figure 20 reports the detail of the distribution of disclosed patents for EPO and USPTO disclosed basis patents with at most 50 claims (95 percent of the disclosed patents). Interestingly, we observe a large concentration of disclosed patents that have exactly 20 claims for USPTO patents and exactly 15 claims for EPO patents.

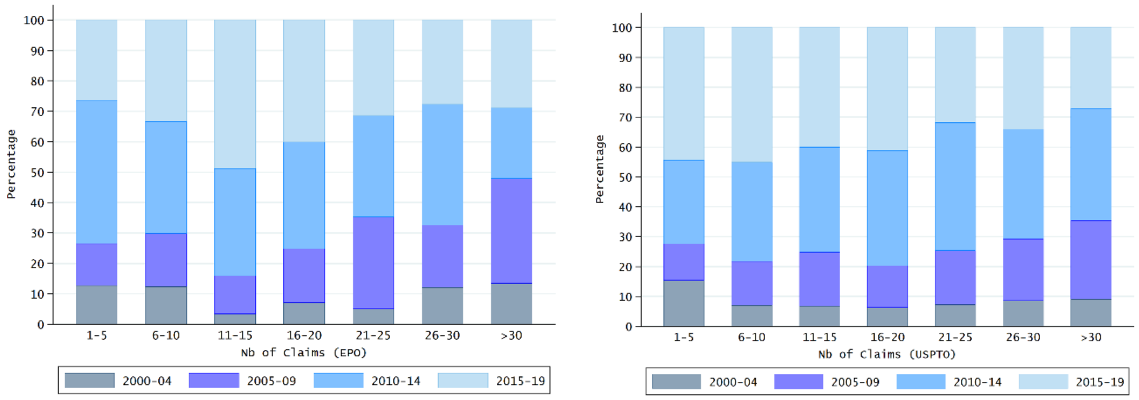
Figure 20. Distribution of disclosed SEPs by number of claims for EPO (left) and USPTO (right) disclosed basis patents.



Note: the two panels are based on different subsets. In the set of EPO basis patents, disclosed patents filed by European firms may be prevailing, whereas in the set of USPTO, disclosed patents filed by U.S. firms may be prevailing. As a consequence, the two sets therefore may not be comparable.

Figure 21 reports instead the share of disclosed patents as divided into seven categories based on their claims number, by the timing of disclosure. This figure suggests that the number of claims per disclosed patents is pretty stable over time both for EPO and USPTO disclosed basis patents.

Figure 21. Share of disclosed basis patents by number of claims and time of disclosure for EPO (left panel) and USPTO (right)

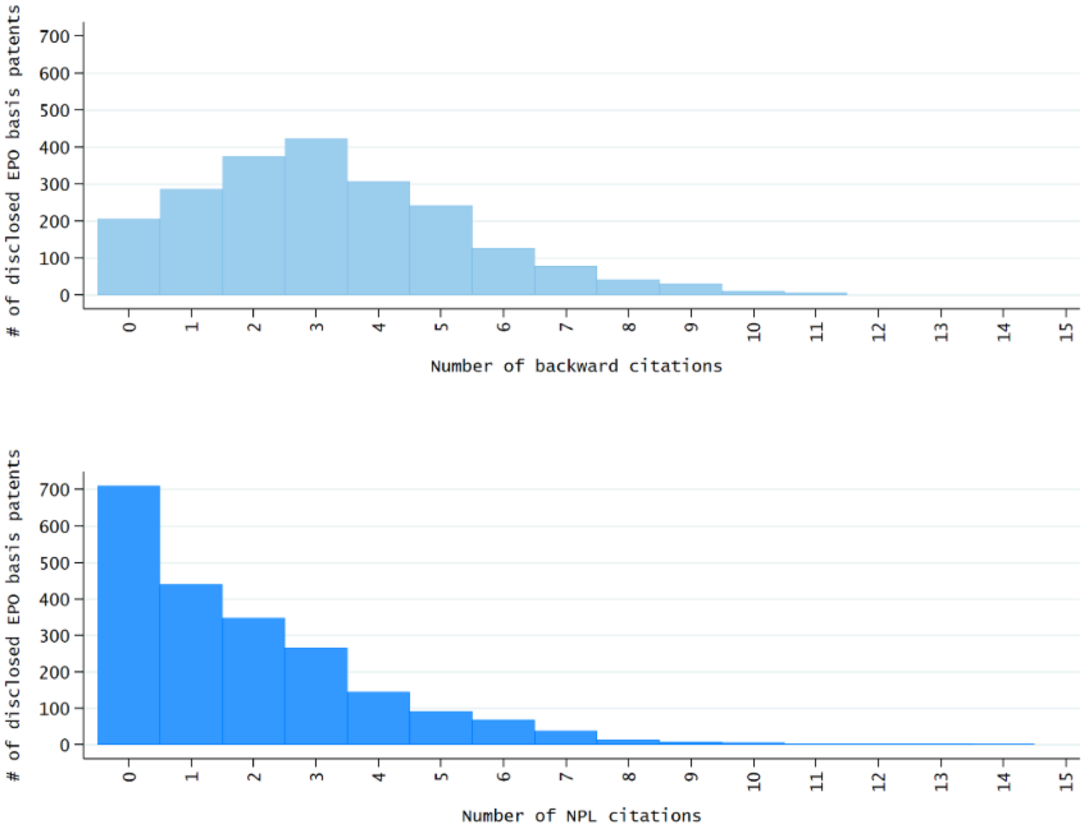


[52] Note: claims are only available for 9,608 patents in our sample, granted USPTO and EPO patents).

As discussed in Section 3.1, the variables *Nb. of backward citations* and *Nb. of non-patent literature* count the number of references to the prior art made by a patent. From Table 8 (on page 23) it is clear that both these variables have a skewed distribution: on average disclosed SEPs make about 10 references to earlier patents, whereas the median is four.

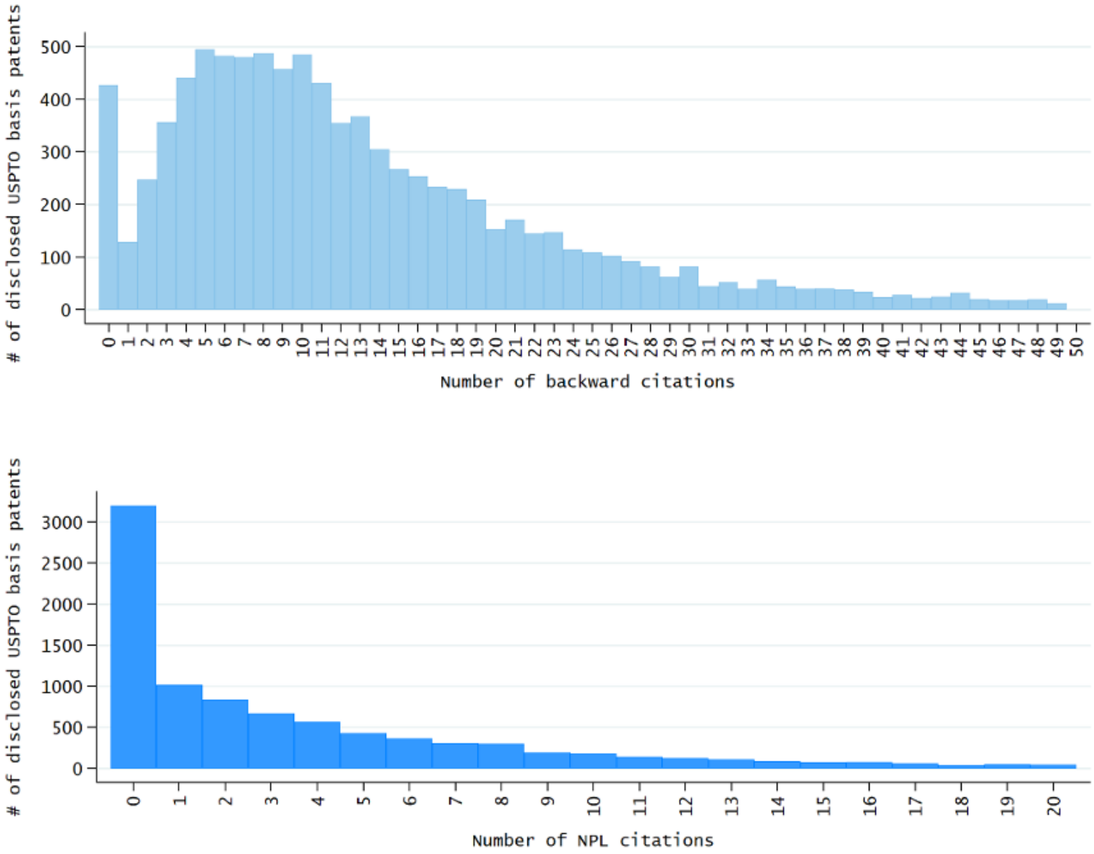
Figure 22 and Figure 23 illustrate the distributions of backward citations and non-patent literature references for EPO and USPTO disclosed basis patents, respectively. Our data confirm the common understanding that the number of backward citations and references to the non-patent literature made by a patent depend to a significant degree on the institutional rules of the patent office where the application is filed. For instance, applicants filing a patent application at the USPTO have to comply with the so-called *duty of candour*. This rule requires applicants to disclose to the patent office all information known to them to be material to patentability. Such a requirement results in a larger number of backward citations and references to the non-patent literature, compared to applications filed at other patent offices, such as the EPO. Indeed, we see that the median number of backward citations for USPTO is 11, whereas that for the EPO is 3. For NPL citations, the median of USPTO patents is two and for the EPO is one. Nevertheless, among both EPO and USPTO disclosed basis patents there is a similar proportion of patents, about one third, that makes no reference at all to the non-patent literature.

Figure 22. Distribution of disclosed patents by number of backward and NPL citations for EPO disclosed basis patents.



We observe some extreme outliers in these variables. One of the disclosed SEP in our sample listed more than a thousand patents as relevant prior art. We see a similar pattern for the references to the non-patent literature. On average, disclosed SEPs make 3.7 references to the non-patent literature, but the median value (only one reference) is considerably smaller than the mean. We also observe that very few disclosed SEPs (5) make more than a hundred references to the non-patent literature.

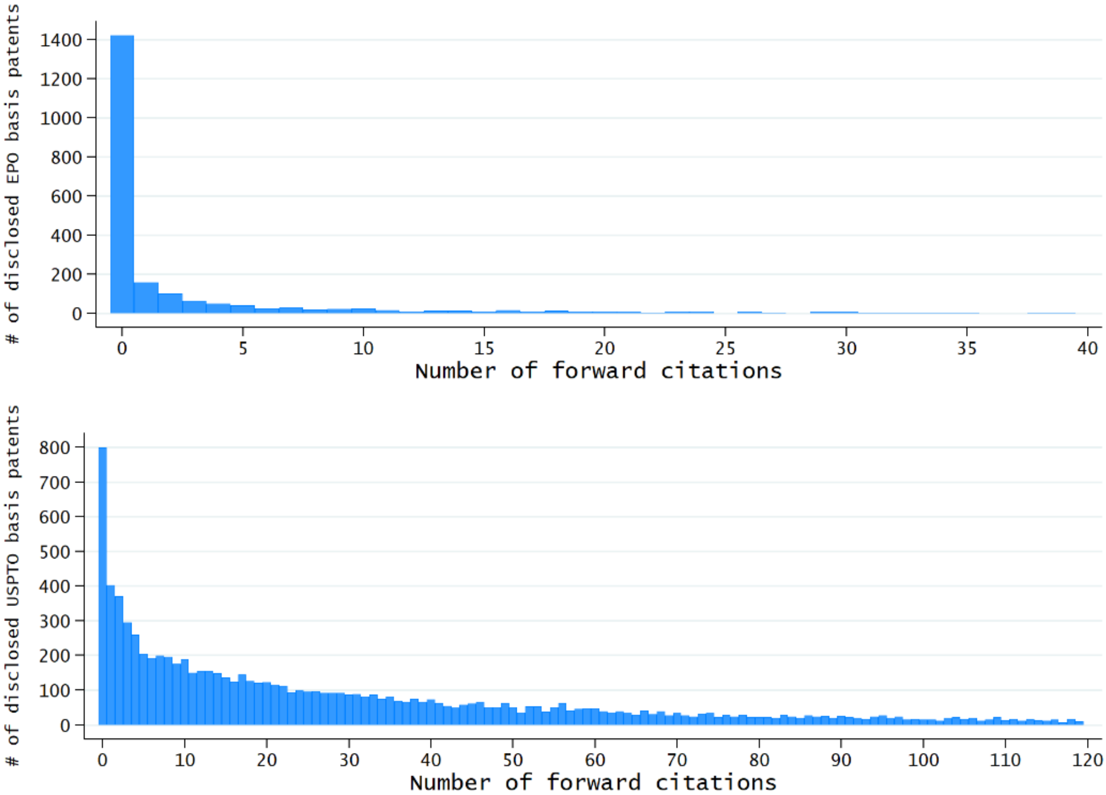
Figure 23. Distribution of disclosed patents by number of backward and NPL citations for USPTO disclosed basis patents.



Also the distribution of the variable *Nb. of forward citations* is heavily skewed. Table 8 (on page 23) reported that on average disclosed SEPs receive about 24 citations from later patents. Nevertheless, the median in our sample is only four forward citations, with a few patents (5) collecting more than a thousand forward citations each. A considerable share of disclosed SEPs (37 percent) did not receive any forward citation so far. Almost 70 percent of the disclosed patents received less than ten forward citations, and 90 percent of the disclosed SEPs less than 50 forward citations. However as discussed above, these figures are heavily affected by the institutional framework of the receiving office.

Figure 24 displays the distribution of disclosed patents by the number of forward citations received for the EPO and USPTO disclosed basis patents. As expected, USPTO disclosed basis patents receive substantially more citations than their EPO counterparts. The average number of forward citations received by a USPTO disclosed basis patent is 44, whereas EPO disclosed basis patent on average receives only 4 citations. Interestingly, a relatively large share of EPO disclosed patents (66 percent) received no forward citation so far, whereas only 8 percent of USPTO disclosed basis patents has not received any citation yet.

Figure 24. Distribution of disclosed EPO (above) and USPTO (below) basis patents by number of forward citations.



3.8 Legal status

Finally, Table 8 (on page 23) already reported information about the variable *Granted,* which reports whether the basis patents (or applications, see above) disclosed to ETSI were actually granted patent protection in the jurisdiction where they were filed.[53] Note that when a patent is not granted, it may be rejected, but also may be waiting for a decision (pending), withdrawn or abandoned by its applicant. On average, the disclosed SEPs (basis patent) obtained patent protection in 73 percent of the cases. However, this statistic is clearly affected by the duration of the prosecution process, which, at several patent authorities, could take on average more than three years. If we considered only patents applied for before the year 2015, we would observe a much higher grant rate, close to 87 percent. Figure 25 illustrates well the importance of this effect of data truncation, with the overwhelming majority of disclosed SEPs being granted patent protection before 2015, and a falling share of granted patents towards the end of the period taken into account.

Figure 26 reports the grant rate by the main patent authorities for disclosed SEPs filed before 2015. As the chart shows, all the main patent offices granted more than 70 percent of the disclosed SEP applications. However, some offices like the USPTO, the Korean and the Japanese patent office have a substantially higher grant rate than the European and the Chinese patent office.

53 As discussed above, we did not consider international applications (WO/PCT) here, as these applications cannot be granted as such.

Figure 25. Distribution of disclosed SEPs by legal status and application year (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).

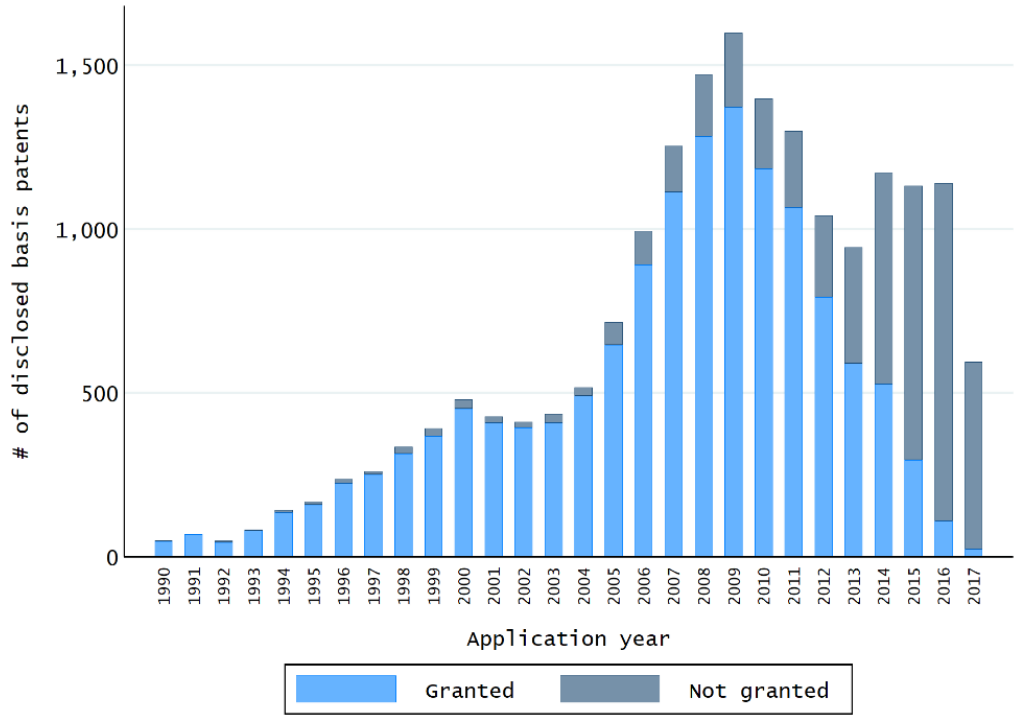
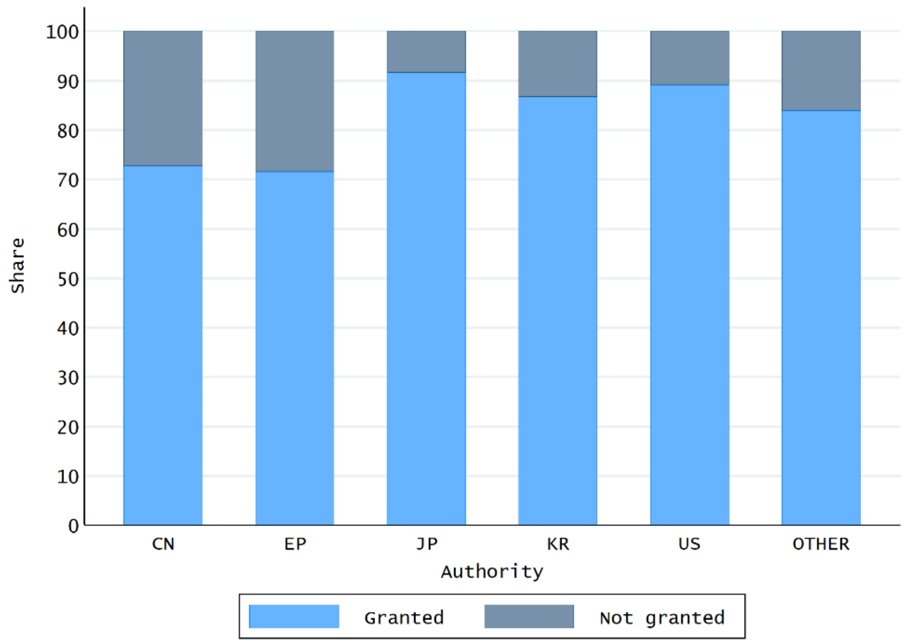


Figure 26. Grant rate by main patent authorities for disclosed SEPs filed before 2015 (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).



4 Disclosed SEPs quality assessment

As part of this project, we were asked by the European Commission to investigate the quality of disclosed SEPs. In this section we aim at assessing how the quality of disclosed SEPs compare to a group of suitable control patents (i.e., never disclosed to ETSI).

The term 'quality' can mean very different things to different people; therefore, we first discuss what type of characteristics we look at in this chapter in Section 4.1. In Section 4.2 we discuss the sample of disclosed patents we will use for the analysis, in Section 4.3 the variables we use, and in 4.4 explain how we created the control set. Section 4.5 presents the analysis.

4.1 Patent quality: technical merit and economic value

As indicated, the term 'quality' can mean very different things to different people. One can, at least, think of the following 'dimensions' of quality:

- **technical merit**: the technological contribution to the field; could be described as the "size of the technological inventive step" associated with the invention that eventually receives a grant, as compared to the prior art.
- **economic value**: economic returns that a patent generates, in terms of direct revenue, or in terms of other benefits from having an exclusive right (e.g. market position for a product). The latter category is sometimes coined the 'patent premium'.
- **legal quality**: the likelihood that a patent survives if challenged in court.

In this study we will focus on the first two categories. The challenge here is that none of the above dimensions can be directly measured. At best, one can look at observable characteristics of patents that are believed to be associated with these dimensions, which are often referred to as proxies (see Squicciarini et al., 2013, for an overview). The second challenge is that these dimensions are related to each other: a patent with a high technical merit is quite likely to be a patent with high economic value (although that does not always need to be true). Indeed, in the literature, technical merit and economic value are found to be positively correlated (de Rassenfosse and Jaffe, 2014). Relatedly, some observable characteristics are associated with both dimensions. A third challenge is that proxies are indirect and per definition imperfect. Studies have shown that, even when a number of proxies are combined in an optimum way, there is still a significant degree of unexplained variance in measuring a dimension like patent value (Gambardella et al., 2008).

Table 11. Observable patent characteristics related to different quality dimensions

Observable characteristics	Can be measured as	Considered as a proxy of technical merit	Considered as a proxy of economic value
Forwards citations	Total number of citations received from any other patent publication (see §3.7).	Yes	Yes
Number of legislations	Number of legislations covered by the applications in the same ETSI patent family (see §3.7).		Yes
Number of claims	Maximum number of claims reported by the publication documents (see §3.7).	Yes	
Years of renewal	Number of years for which the patent is renewed (see §4.3).		Yes

The discussion in the economic literature on the pros and cons on the various proxies is complex and lively, and it is outside the scope of this study offer a voluminous treaty on this topic, and we will follow a pragmatic approach. We focus on a selection of widely established metrics. Table 11 introduces the observable characteristics we will focus on, and how they are related to technical merit and economic value. In addition, we will also look at an existing composite measurement as used by the OECD, associated with both dimensions.

4.2 Mapping ETSI disclosed patents to EPO/USPTO databases

The disclosed basis SEPs discussed in Chapter 2 and 3 come from many different legislations. Yet, for many of the observable patent characteristics we consider in this chapter, it is known that they differ considerably across legislations. For instance, a patent at the USPTO has a much higher likelihood of receiving citations than one at the EPO. Therefore, patents from a given office should be only compared to other patents from that same office. Because the analysis may to some degree also depend on specific (institutional) aspects of the patent office,[54] we decided not to look at patents from one office only but look at patents from two offices: EPO and USPTO. For reasons of comparability, we furthermore limit our analysis to granted patents.

Our set of EPO patents for this chapter consists of the granted, disclosed basis EPO patents, plus granted EPO patents that are family member of a disclosed base patent of another patent office.[55] Likewise, our set of USPTO patents for this chapter consists of the granted, disclosed basis USPTO patents, plus granted USPTO patents that are family member of a disclosed base patent of another patent office.

Our initial EPO set includes 4,758 granted patents, and our final USPTO set includes 13,230 granted patents. Both sets together cover 17,994 granted patents. (Later we will slightly reduce the number of patents in these sets when we are creating the control samples.)

4.3 Observable characteristics related to patent quality

For this study, we focus on a set of observable characteristics (variables) that are generally related to the quality of patents. These variables, together with their sources, are listed in Table 12. Variables calculated from PATSTAT are already described in detail in Sections 0 and 3.7. Below, we discuss the variables that were not already presented in that section.

The variable RENEWALS is the number of years for which the patent is active starting from the filing year.[56] Patents rights generally last up to 20 years; however, they need to be periodically renewed. As renewal is costly, and some offices use increasing fees, applicants will be willing to pay an increasing amount of money only for those patents that are valuable in their own perception (Lanjouw et al., 1998). Similar to the number of legislations and the number of claims, renewals are directly related to the private costs of protection that an applicant is willing to pay (Pakes and Schankerman; 1984).

As there are numerous indicators linked to different aspects of patent quality, there were some attempts to construct a composite index. Lanjouw and Shankerman (2004) use a multiple-indicator model with one common latent factor. This latent factor is the unobserved characteristic of a patented innovation that simultaneously influences the number of forward citations, the number of backward citations, the number of claims, and the number of legislations. Following this approach, the variable QUALITY_INDEX results from a latent variable model, where the components are as follows: the number of forward citations (up to 5 years after publication), INPADOC patent family size,[57] number of claims, and the patent generality

54 See for instance Figure 20, where one can see different peaks for the two patent offices, related to the different fee structure used by these offices.

55 As the ETSI family might include more than one EPO patent, we choose the one with the earliest filing date. In the case of multiple patents filed on the same filing date, we randomly select one of them.

56 Definition taken from Squicciarini et al. (2013), where the OECD data source is described we use for this variable.

57 Note that OECD Quality Database calculates this composite quality index using the INPADOC family definition and not on the ETSI family definition used in other analysis presented in Section 4.

index[58]. Such variable then captures the technological importance of the invention and its market opportunities.

Table 12. Description of the variables used in the value analysis and their source

Variable	Definition	Data source
NUM_FORW_CIT	Total number of citations received from any other patent publication (see §3.7).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)
NUM_LEGISLATIONS	Number of legislations covered by the applications in the same ETSI patent family (see §3.7).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)
NUM_CLAIMS	Maximum number of claims reported by the publication documents (see §3.7).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)
RENEWALS	Number of years for which the patent is renewed (see §4.3).	OECD Quality Database (Version 2019)
QUALITY_INDEX	Composite index calculated using the number of forward citations (up to 5 years after publication), INPADOC family size, number of claims and the patent generality index (see §4.3).	OECD Quality Database (Version 2019)
NUM_BACK_CIT	Number of citations to patents made by the patent (see §3.7).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)
NUM_NPL_CIT	Number of citations to scientific articles made by the patent (see §3.7).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)
NUM_IPC_4_DIGIT	Number of distinct IPC classes (4 digit) reported in the patent (see §0).	Our calculations using the EPO-PATSTAT Database (Autumn, 2018)

4.4 The construction of the control set

Now we have identified the disclosed patents to be used in the analysis (Section 4.1), we need to identify suitable control patents. Following a standard approach in the literature (Rysman and Simcoe, 2008; Bekkers et al. 2016), the control is built by matching the patent office, the filing year, and the IPC main group.[59] Furthermore, to control for potential differences in the application procedure, we also match patents on whether or not there is an earlier Patent Convention Treaty (PCT) application of the patent. The data source for the control set is the EPO-PATSTAT Database.

Furthermore, to prevent that selected control patents are actually a family member of a disclosed ETSI patent, we ensure that none of the control patents is part of the INPADOC patent family of any ETSI disclosed patent.[60] In Figure 27, we illustrate this graphically (for the same patents used in Figure 6 and Figure 7 of Section 2.6, above).

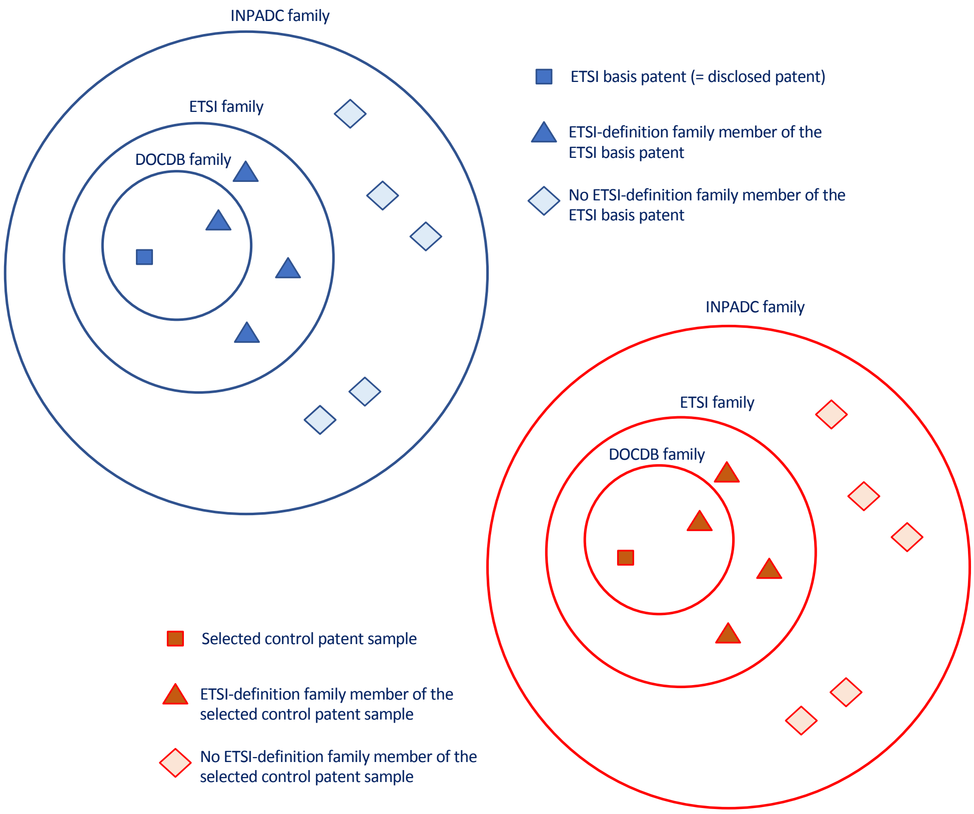
Depending on whether a patent is granted at the USPTO or the EPO, the matching procedure differs. USPTO patents have a primary class that makes straightforward the choice on which IPC main group to perform the matching. On the contrary, EPO patents report IPC classes in alphabetical order, so we decided to assign each application to the most recurrent IPC main group. A similar approach is used for 1,509 USPTO patents that do not have a primary IPC main group.

58 The generality index captures the range of technological field that are affected by an invention (Henderson et al., 1998). This indicator is based on the Hirschman-Herfindahl Index (HHI) of the dispersion of the distribution of the shares of forward citations over the IPC classes.

59 PC classification symbols are made up of a letter denoting the IPC section (e.g. A), followed by a number (two digits) denoting the IPC class (e.g. A63), then a letter denoting the IPC subclass (e.g. A63B). A number (variable, 1-3 digits) denotes the IPC main group (e.g. A63B49). This is followed by a forward slash "/" and a number (variable, 1-3 digits) denoting the IPC subgroup (e.g. A63B49/02).

60 While technically we could also have decided that the two patent families following the ETSI family definition should not overlap, this would require us to calculate ETSI-definition family members for all potential control group member patents. With our simpler, INPADOC non-overlapping approach, our goal is already reached as well. After all, if the control group patent is not in any INPADOC family of an ETSI patent, it can be definition never be member of an ETSI-defined family of an ETSI patent.

Figure 27. ETSI disclosed patents and control group patents



From the pool of potential controls, containing over 5.8 million patents, we randomly sampled – without replacement – up to 5 control patents for each disclosed patent. For a (very) small fraction of the ETSI disclosed patents, no matching control patents could be identified. For this reason, we removed these disclosed patents from our analysis.

Ultimately, our EPO set consists of 4,607 granted disclosed EPO patents and 19,477 matched control patents, and our USPTO set consists of 12,832 granted disclosed USPTO patents and 56,100 matched control patents.[61]

[61] On average each EPO and USPTO disclosed patent has 4.22 and 4.37 control patents respectively.

4.5 Analysis if disclosed vs. control group patents

First, we compare the sample means of the selected variables for the disclosed patents and the control group. The results are shown in Table 13 and Table 14.

Table 13. Test for equality of the mean (disclosed vs. controls) for the EPO patents

	Controls		Disclosed			
Variable	**Obs.**	**Mean**	**Obs.**	**Mean**	**Diff**	**t-stat**
NUM_FORW_CIT	19,477	1.812	4,607	2.986	-1.174	-10.549***
NUM_LEGISLATIONS	19,477	6.145	4607	8.201	-2.056	-38.862***
NUM_CLAIMS	19,477	15.370	4,607	17.608	-2.239	-13.798***
RENEWAL	19,477	9.995	4,607	10.786	-0.790	-11.145***
QUALITY_INDEX	19,475	0.234	4,605	0.287	-0.052	-28.542***
NUM_BACK_CIT	19,477	4.020	4,607	3.585	0.436	9.063***
NUM_NPL_CIT	19,477	1.395	4,607	2.252	-0.857	-21.666***
NUM_IPC_4_DIGIT	19,477	1.730	4,607	1.856	-0.126	-8.212***

Legend: * p<0.05, ** p<0.01, *** p<0.001

Table 14. Test for equality of the mean (disclosed vs. controls) for the USPTO patents

	Controls		Disclosed			
Variable	**Obs.**	**Mean**	**Obs.**	**Mean**	**Diff**	**t-stat**
NUM_FORW_CIT	56,100	22.296	12,832	41.202	-18.906	-42.309***
NUM_LEGISLATIONS	56100	4.029	12832	7.241	-3.212	-100.077***
NUM_CLAIMS	56,100	18.249	12,832	20.849	-2.600	-20.969***
RENEWAL	56,100	7.162	11,473	7.676	-0.514	-12.251***
QUALITY_INDEX	56,095	0.282	11,473	0.359	-0.077	-62.784***
NUM_BACK_CIT	56,100	19.856	12,832	18.907	0.949	3.313***
NUM_NPL_CIT	56,100	5.533	12,832	7.591	-2.058	-15.949***
NUM_IPC_4_DIGIT	56,100	2.025	12,832	2.137	-0.112	-10.786***

**Legend: *** p<0.05, ** p<0.01, *** p<0.001

We observe two important things. Firstly, we see that for most variables, the sample means for the USPTO are different (often higher) than those for the EPO. This is related to institutional differences between these offices, like differences in the working procedures of the examiners, or the duties of applicants.[62] Secondly, within each of these offices, we see clear differences between disclosed patents and controls patents. In fact, for all the metrics used, except backward citations, disclosed patents have statistically significant higher average values. For instance, EPO disclosed patents receive on average 3.0 citations, whereas the EPO control group score of 1.8. Similarly, the USPTO disclosed patents receive on average

[62] A good example here is the large difference in magnitude for the variables NUM_BACK_CIT and NUM_NPL_CIT. This is explained by the duty of candor at the USPTO. This duty obliges applicants at the USPTO to a more binding commitment in disclosing prior art, resulting in a larger number of backward citations reported in the USPTO document as compared to the EPO, and hence, by on average more received forward citations.

41.2 citations, whereas the USPTO control group receives 22.3. Both results suggest that disclosed patents have both a higher technical merit[63] and a higher economic value.

Regarding the number of countries in which applicants seek protection (variable NUM_LEGISLATIONS), EPO disclosed patents have on average a larger geographical scope as they are applied in 8.2 legislations, whereas the control group this is 6.1. We see a similar difference for the USPTO patents. Furthermore, both technological breadth measured using patent claims (NUM_CLAIMS) and length of patent validity measured with renewals (RENEWAL) are higher for disclosed patents than controls, both at the EPO and the USPTO. Finally, the composite index of patent quality we use (QUALITY_INDEX) also indicates that EPO and USPTO disclosed patents have higher quality than the control set.

Backward citations (NUM_BACK_CIT) is the only variable where we see a lower score for disclosed patents compared with the control patents. While this result might look counterintuitive, it might be due to some form of strategic citing. Lampe (2012) shows that applicants with large patent portfolios in sectors such as computer and electronics strategically retain more known prior-art. Even if this study is not specifically related to standard-essential patents, we cannot exclude that this result on backward citations might be driven by strategic withholding.

The above observations for differences between the averages is interesting, but taken in isolation, they may not tell the full story. Therefore, we also do a multivariate regression analysis, to control simultaneously for multiple patent characteristics. The regression specification we use is:

$$Y_i = f(\alpha + \beta DECLARED_i + \gamma X_i + \lambda_t + \theta_c + \varepsilon_i) \qquad (1)$$

where Y_i is a measure of quality for patent i as: number of forward citations (*NUM_FORW_CIT*), number of legislations (*NUM_LEGISLATIONS*), number of claims (*NUM_CLAIMS*), number of years the patent is renewed (*RENEWAL*), and the composite index of quality (QUALITY_INDEX). The variable of primary interest for this analysis is dummy variable *DISCLOSED* that is equal to one if patent i is disclosed essential to ETSI or there is a disclosed essential patent in its family (according to the ETSI definition), and 0 otherwise. Vector X_i is a set of control variables that include number of backward citations (*NUM_BACK_CIT*), number of citations to previous scientific literature (*NUM_NPL_CIT*), number of 4-digit IPC classes reported in the patent (*NUM_IPC_4_DIGIT*), and a dummy variable equal to one if there is an earlier PCT international application (*INTERNATIONAL_APPL*). Finally, λ_t and θ_c are a set of filing year and technology class fixed-effects, respectively. To unfold possible differences in quality that might emerge from institutional differences in the prosecution process, all these models are separately estimated for the EPO and USPTO patents.

[63] Concerning technical merit, it is important to acknowledge that forward citations performance may also be a result of the fact that a patent is disclosed as potentially essential. Concerning economic value, this possible bias is less relevant, as the fact that a patent is potentially essential in fact can increase its economic value.

We estimate Equation 1 using an Ordinary Least Squares (OLS) model. However, as some of the dependent variables are count variables, we also employ a Poisson Quasi Maximum Likelihood (PQML) model to correct for potential overdispersion in our dependent variables.

Table 15 and Table 16 report the descriptive statistics of the variables used in the regression analysis for the EPO and USPTO, respectively.

Table 15. Summary statistics of the variables for the EPO patents

Variable	Observations	Mean	Std. Dev.	Min	Max
NUM_FORW_CIT	24,084	2.037	6.808	0	194
NUM_LEGISLATIONS	24,084	6.538	3.329	1	39
RENEWAL	24,084	10.147	4.338	1	21
QUALITY_INDEX	24,080	0.244	0.114	0.038	0.877
DISCLOSED	24,084	0.191	0.393	0	1
NUM_BACK_CIT	24,084	3.937	2.938	0	198
NUM_NPL_CIT	24,084	1.559	2.438	0	123
NUM_IPC_4_DIGIT	24,084	1.754	0.940	1	20
NUM_CLAIMS	24,084	15.798	9.942	0	255
INTERNATIONAL_APPL	24,084	0.739	0.439	0	1

Table 16. Summary statistics of the variables for the USPTO patents

Variable	Observations	Mean	Std. Dev.	Min	Max
NUM_FORW_CIT	68,932	25.815	46.254	0	2478
NUM_LEGISLATIONS	68,932	4.627	3.511	1	43
RENEWAL	67,573	7.249	4.098	0	21
QUALITY_INDEX	67,568	0.296	0.123	0.022	0.949
DISCLOSED	68,932	0.186	0.389	0	1
NUM_BACK_CIT	68,932	19.679	29.261	0	3072
NUM_NPL_CIT	68,932	5.916	13.209	0	917
NUM_IPC_4_DIGIT	68,932	2.046	1.064	1	30
NUM_CLAIMS	68,932	18.733	12.712	0	596
INTERNATIONAL_APPL	68,932	0.365	0.482	0	1

Table 17 reports the estimates of our main coefficient of interest, β, in the case of forward citations. For all the estimation methods (OLS and PQML), the variable *DISCLOSED* is positive and highly significant. However, the magnitude of the coefficients of the two different methods cannot be directly compared, as they capture different aspects. In case of a positive sign, the OLS coefficient indicates the expected number of citations additionally received by disclosed patents. The coefficients column show that EPO disclosed patents have additionally 1.2 more citations than the EPO controls patents; whereas USPTO disclosed patents have additionally 19.6 more citations than the USPTO control patents (OLS). These numbers are consistent with the differences reported in Table 13 and Table 14. Poisson coefficients, in turn, can be translated into a percentage change by exponentiating and subtracting one. PQML coefficients are closer in magnitude, and indeed, while the effect is always larger at the USPTO, the percentage increases are more comparable being 68.9 percent for EPO and 88.9 percent for the USPTO.[64]

Table 17. Number of forward citations regressions

Dependent variable: NUM_FORW_CIT				
	EPO patents		**USPTO patents**	
	OLS	**PQML**	**OLS**	**PQML**
DISCLOSED	1.175*** (0.14)	0.518*** (0.05)	19.628*** (0.55)	0.636*** (0.01)
NUM_BACK_CIT	0.107*** (0.02)	0.064*** (0.01)	0.062*** (0.02)	0.001* (0.00)
NUM_NPL_CIT	0.086*** (0.02)	0.033*** (0.01)	0.202*** (0.02)	0.008*** (0.00)
NUM_IPC_4_DIGIT	0.225*** (0.06)	0.071*** (0.02)	1.851*** (0.19)	0.054*** (0.01)
INTERNATIONAL_APPL	-4.747*** (0.14)	-1.997*** (0.04)	-8.595*** (0.27)	-0.433*** (0.01)
CONSTANT	-10.201*** (1.53)	-2.015** (0.87)	71.137 (.)	3.783*** (0.63)
FILING YEAR DUMMY	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES
Observations	24084	24084	68932	68932
Log-Likelihood		-66125.8		-1009494.9
R-squared	0.178		0.226	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

[64] These figures are obtained as exp(0.518)-1 and as exp(0.636)-1.

Table 18 presents the results when the outcome variable is the number of legislations. Also in this case, the coefficient of the variable DISCLOSED is positive and significant, indicating that disclosed patents are expected to have a larger geographic scope. In particular, ETSI families of EPO disclosed patents include 2.0 more legislations than controls do, ETSI families of USPTO disclosed patents include 3.1 more legislations than controls do (OLS). The PQML outcomes confirm the results and indicate that ETSI family members of EPO and USPTO disclosed patents are filed in 32% and 74% more legislations than their controls, respectively.

Table 18. Number of legislations regressions

Dependent variable: NUM_LEGISLATIONS				
	EPO patents		**USPTO patents**	
	OLS	**PQML**	**OLS**	**PQML**
DISCLOSED	2.001***	0.280***	3.061***	0.556***
	(0.06)	(0.01)	(0.04)	(0.01)
NUM_BACK_CIT	-0.062***	-0.009***	0.009***	0.000
	(0.01)	(0.00)	(0.00)	(0.00)
NUM_NPL_CIT	-0.019**	-0.002	0.023***	0.005***
	(0.01)	(0.00)	(0.00)	(0.00)
NUM_IPC_4_DIGIT	0.321***	0.045***	0.360***	0.071***
	(0.03)	(0.00)	(0.02)	(0.00)
INTERNATIONAL_APPL	1.552***	0.244***	1.540***	0.327***
	(0.05)	(0.01)	(0.03)	(0.01)
CONSTANT	5.314***	2.033***	4.168	1.931***
	(1.63)	(0.20)	(1001.06)	(0.33)
FILING YEAR DUMMY	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES
Observations	24084	24084	68932	68932
Log-Likelihood		-57658.5		-170361.4
R-squared	0.205		0.228	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

Table 19 shows the results for the technological scope of patents as measured by the number of claims. Also in this case, the variable DISCLOSED is positive and significant in all the estimated models. EPO disclosed patents have 2.4 additional claims (OLS), which is 15.6% more claims (PQML) than the patents in the control set. USPTO disclosed patents have 2.9 additional claims (OLS), which is 15.8% more claims (PQML) than the patents in the control set.

Table 19. Number of claims regressions

Dependent variable: NUM_CLAIMS				
	EPO patents		**USPTO patents**	
	OLS	**PQML**	**OLS**	**PQML**
DISCLOSED	2.394***	0.145***	2.868***	0.147***
	(0.20)	(0.01)	(0.14)	(0.01)
NUM_BACK_CIT	0.273***	0.016***	0.037***	0.001***
	(0.04)	(0.00)	(0.01)	(0.00)
NUM_NPL_CIT	0.014	0.001	0.037***	0.002***
	(0.03)	(0.00)	(0.01)	(0.00)
NUM_IPC_4_DIGIT	0.036	0.003	0.055	0.006**
	(0.07)	(0.00)	(0.05)	(0.00)
INTERNATIONAL_APPL	-0.278*	-0.017*	-4.793***	-0.273***
	(0.15)	(0.01)	(0.10)	(0.01)
CONSTANT	3.134	1.867***	8.498	2.213***
	(2.75)	(0.38)	(.)	(0.26)
FILING YEAR DUMMY	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES
Observations	24084	24084	68932	68932
Log-Likelihood		-108301.4		-370286.3
R-squared	0.0507		0.0728	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

Table 20 reports the estimates for the model where the outcome is the number of years in which the patent is valid on the basis of paid renewal fees. The coefficient of the variable *DISCLOSED* is positive and significant for all the specifications suggesting that disclosed patents have a longer active life. In particular, EPO disclosed patents are expected to last 0.94 year longer (OLS), which corresponds to a 9.5 percent increase (PQML), as compared to controls. Regarding the USPTO disclosed patents, they are expected to last 0.26 year longer (OLS), that corresponds to a 3% increase, as compared to controls (PQML).

Table 20. Number of renewals regressions

Dependent variable: RENEWAL				
	EPO patents		**USPTO patents**	
	OLS	**PQML**	**OLS**	**PQML**
DISCLOSED	0.944***	0.091***	0.225***	0.030***
	(0.03)	(0.00)	(0.01)	(0.00)
NUM_BACK_CIT	0.029***	0.003***	0.001***	0.000***
	(0.01)0	(0.00)	(0.00)	(0.00)
NUM_NPL_CIT	0.000	0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)
NUM_IPC_4_DIGIT	0.118***	0.010***	-0.010*	-0.002**
	(0.02)	(0.00)	(0.01)	(0.00)
INTERNATIONAL_APPL	0.059	0.006*	0.128***	0.021***
	(0.04)	(0.00)	(0.01)	(0.00)
CONSTANT	8.429***	2.375***	2.444	0.609***
	(2.30)	(0.16)	(441.22)	(0.19)
FILING YEAR DUMMY	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES
Observations	24084	24084	67573	67573
Log-Likelihood		-53712.3		-132146.5
R-squared	0.753		0.891	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

Table 21 reports the estimates for the model where the outcome is the OECD composite quality index indicator. While all previous dependent variables were count variables, this index indicator is a continuous variable. Therefore, we only estimate the OLS model. Furthermore, this indicator does not have a scale, so the interpretation of the coefficient is not straightforward. Still, the positive and significant coefficient indicates that both EPO and USPTO disclosed patents have higher quality index than the controls patents.

Table 21. Quality composite index regressions

Dependent variable: QUALITY_INDEX		
	EPO patents	**USPTO patents**
	OLS	**OLS**
DISCLOSED	0.050***	0.071***
	(0.00)	(0.00)
NUM_BACK_CIT	-0.001***	0.001***
	(0.00)	(0.00)
NUM_NPL_CIT	-0.000	0.001***
	(0.00)	(0.00)
NUM_IPC_4_DIGIT	0.017***	0.020***
	(0.00)	(0.00)
INTERNATIONAL_APPL	0.004**	-0.016***
	(0.00)	(0.00)
CONSTANT	0.315***	0.255***
	(0.06)	(0.07)
FILING YEAR DUMMY	YES	YES
IPC DUMMY	YES	YES
Observations	24080	67568
R-squared	0.163	0.247

Note: OLS estimations with clustered standard errors reported in parenthesis. Legend: * p<0.05, ** p<0.01, *** p<0.001

So far, our analyses have indicated that, irrespectively of the outcome variable used, disclosed patents score higher (i.e., are of higher quality) than control patents. In the remaining part of the report we further explore the presence of possible temporal patterns. Table 22 and Table 23 report the estimates for the number of forward citations for EPO and USPTO disclosed patents respectively, over three time-period starting in 1990.[65] The number of additional expected citations for both the EPO and USPTO disclosed patents decrease over time from 2.89 to 0.21 and from 47.77 to 8.94, respectively. However, this decreasing trend might occur because patents recently filed have less time to accumulate citations and they might display a lower average number of citations. The PQML result seems to confirm this interpretation for the USPTO as its coefficient increases in the last period. The USPTO patents filed between 1990 and 1999 receive about 88.7% more forward citations than the random controls; whereas this percentage increases up to 103% for the USPTO patents filed between 2010 and 2017. However, for the EPO, the PQML estimation confirms the decrease also in terms of percentages. The EPO patents filed between 1990 and 1999 receive about 57.8% more forward citations than the random controls; whereas this percentage decrease to 46% for the EPO patents filed between 2010 and 2017.

Table 22. Number of forward citations regressions over time for EPO

Dependent variable: NUM_FORW_CIT						
	1990-1999		**2000-2009**		**2010-2017**	
	OLS	**PQML**	**OLS**	**PQML**	**OLS**	**PQML**
DISCLOSED	2.889***	0.465***	1.220***	0.592***	0.212***	0.376***
	(0.66)	(0.09)	(0.17)	(0.06)	(0.06)	(0.08)
NUM_BACK_CIT	0.158*	0.040***	0.128***	0.076***	0.029***	0.061***
	(0.08)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
NUM_NPL_CIT	0.236	0.036	0.100***	0.032***	0.021***	0.034***
	(0.16)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
NUM_IPC_4_DIGIT	0.498***	0.080***	0.089*	0.042*	0.069*	0.132**
	(0.19)	(0.03)	(0.05)	(0.02)	(0.04)	(0.06)
INTERNATIONAL_APPL	-10.086***	-2.411***	-4.132***	-1.882***	-1.450***	-1.575***
	(0.50)	(0.13)	(0.15)	(0.05)	(0.12)	(0.08)
CONSTANT	7.985***	1.054	4.169***	-13.446	-0.940	-11.309***
	(1.58)	(0.66)	(0.44)	(19.83)	(0.85)	(2.88)
FILING YEAR DUMMY	YES	YES	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES	YES	YES
Observations	3518	3518	13016	13016	7350	7350
Log-Likelihood		-17815.4		-37464.0		-8713.0
R-squared	0.217		0.124		0.101	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

65 The 591 patents filed before 1990 are excluded from the analysis.

Table 23. Number of forward citations regressions over time for USPTO patents

Dependent variable: NUM_FORW_CIT						
	1990-1999		2000-2009		2010-2017	
	OLS	PQML	OLS	PQML	OLS	PQML
DISCLOSED	47.768***	0.635***	22.425***	0.598***	8.938***	0.711***
	(3.73)	(0.04)	(0.76)	(0.02)	(0.34)	(0.02)
NUM_BACK_CIT	0.602***	0.006***	0.114***	0.002***	-0.013	-0.002***
	(0.14)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)
NUM_NPL_CIT	1.031**	0.007**	0.260***	0.006***	0.163***	0.011***
	(0.45)	(0.00)	(0.03)	(0.00)	(0.02)	(0.00)
NUM_IPC_4_DIGIT	7.750***	0.107***	1.111***	0.030***	0.423***	0.036***
	(1.18)	(0.01)	(0.22)	(0.01)	(0.12)	(0.01)
INTERNATIONAL_APPL	-21.013***	-0.430***	-10.458***	-0.386***	-5.362***	-0.526***
	(2.42)	(0.05)	(0.45)	(0.02)	(0.22)	(0.02)
CONSTANT	15.020	3.542***	33.271*	3.469***	12.807***	2.384***
	(33.28)	(0.46)	(17.43)	(0.44)	(3.15)	(0.37)
FILING YEAR DUMMY	YES	YES	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES	YES	YES
Observations	6722	6722	32613	32613	29206	29206
Log-Likelihood		-203364.4		-545814.0		-228853.9
R-squared	0.157		0.159		0.160	

Note: OLS estimations with clustered standard errors reported in parenthesis. Poisson Quasi Maximum Likelihood (PQML) estimations with clustered standard errors. Legend: * p<0.05, ** p<0.01, *** p<0.001

The estimations carried out using the OECD composite quality index are reported in Table 24, and indicate contrasting results. For the EPO patents, the coefficient of the variable *DISCLOSED* decreases in the second period and increases again for the period between 2010 and 2017. The last increase is rather small as compared to the decrease, so we can conclude that the overall quality of disclosed EPO patents does not increase over time. The results for the USPTO are more evident and they indicate that on average the quality of USPTO disclosed patents decreased over time.

Table 24. Quality composite index regressions over time for EPO and the USPTO

Dependent variable: QUALITY_INDEX						
	EPO			**USPTO**		
	1990-1999	**2000-2009**	**2010-2017**	**1990-1999**	**2000-2009**	**2010-2017**
	OLS	**OLS**	**OLS**	**OLS**	**OLS**	**OLS**
DISCLOSED	0.080***	0.043***	0.046***	0.112***	0.080***	0.046***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NUM_BACK_CIT	0.004***	-0.001***	-0.001***	0.002***	0.001***	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NUM_NPL_CIT	-0.000	-0.000	0.000	0.001**	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NUM_IPC_4_DIGIT	0.027***	0.018***	0.003*	0.029***	0.017***	0.022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
INTERNATIONAL_APPL	0.013***	0.001	0.003	0.024***	-0.009***	-0.026***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CONSTANT	0.263***	0.177***	0.354***	0.314***	0.301***	0.200***
	(0.05)	(0.01)	(0.07)	(0.07)	(0.06)	(0.03)
FILING YEAR DUMMY	YES	YES	YES	YES	YES	YES
IPC DUMMY	YES	YES	YES	YES	YES	YES
Observations	3518	13016	7350	6713	32549	27915
R-squared	0.192	0.0931	0.296	0.277	0.251	0.177

Note: OLS estimations with clustered standard errors reported in parenthesis. Legend: * p<0.05, ** p<0.01, *** p<0.001

To conclude, all the analyses we carried out to assess the quality of patents disclosed as SEPS relative to the sample of suitable controls, point in the same direction. Disclosed SEPs score higher in terms of all the main proxies for patent quality that are commonly used in the patent literature. This is true for both those variables usually associated with technical merit (the technological contribution to the field) and those variables usually associated with economic value (the economic returns that a patent generates).

All in all, these results confirm the expectation that patents disclosed to ETSI are of greater technological importance and have more market potential than suitable controls.

That disclosed SEPs have a higher technical merit may indicate that SDOs are able to attract promising technologies (to know whether they are able to include attractive technologies, we would need to know which of these are factually essential). Yet, the higher technical merit may also be the result of the act of disclosure as such. Firstly, disclosed patents are more 'visible'. Secondly, for any party investing in R&D in a technical field where standards are important, it is rational to build such R&D upon knowledge already embedded in these standards, instead of on a 'dead track'. This also increases the likelihood that essential patents, and also the wider set of disclosed, potentially essential patents, receive more citations from future patents.

That disclosed SEPs on average have a higher economic value than comparable, not disclosed patents is not surprising. After all, a subset of these patents will become factually essential patents, will then need to be licensed by all parties in the world that implement that standard,[66] and thus have the potential to generate significant licensing revenues (or other benefits such as cross-licensing opportunities). Hence, it is rational for applications of such patents to seek protection in many countries and renew their patent, which are exactly the variables that are associated with higher economic value of a patent.

[66] To be precise: insofar the product implementation is in a product category that requires the essential patent, and it is not only essential for an optional feature that is not implemented in the product implementation.

5 Summary and conclusions

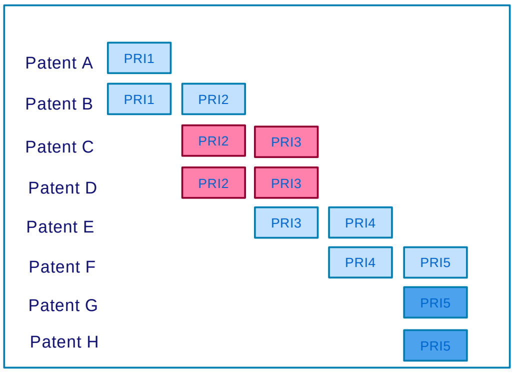
The main goals of this study are (a) to provide a patent landscape analysis of SDO disclosed patents (and what this implies for their use as input to an essentiality assessment mechanism), and (b) to analyse whether SDO disclosed patents differ from comparable other patents in quality (both technical merit and economic value).

Our main conclusions are as follows:

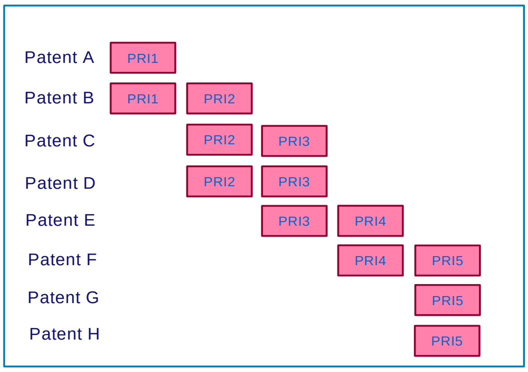
- While the ETSI IPR database of disclosed potentially standard-essential patents is by far the most sophisticated one, it is a non-trivial task to identify patents from that database and clean/harmonize/select/de-duplicate/transform that data into information to be used for a given purpose, such as input for a process of essentiality assessment. Also important is a good understanding of both the ETSI IPR policy, the related procedures, and database aspects. In particular, a good understanding of the meaning of a ‘patent family’ in the ETSI context is important. For good reasons, ETSI uses a tailor-made definition for this concept, and this has significant impact on understanding and interpreting the data for specific purposes. Unlike usual definitions of patent families such as DOCDB and INPADOC, ETSI patent families are ego-families and not mutually exclusive. As a simplified example, Figure 28 shows how a given set of eight patents and their priority relations translate into two DOCDB patent families, one INPADOC family, and eight ETSI patent families (which partly overlap).

  While the ETSI database also contains a large number of ‘non-harmonized records’ (i.e., records that ETSI itself has not yet managed to match with the EPO patent database), we found that these records are not likely to have a large impact on the use of the data.

Figure 28. Differences between patent family definitions, for a given set of patents and their associated priorities (for details and colour coding, see Section 2.2)



Six DOCDB families: one with patent C and D, one with patent G and H, and all other patents a one member family.



One INPADOC family.

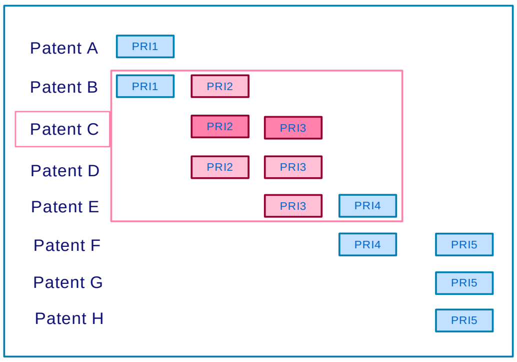
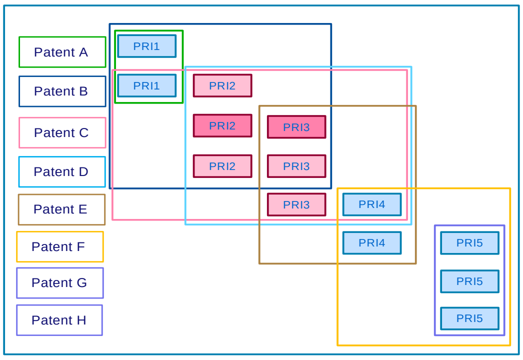


Illustration of the ETSI family definition of the disclosed (‘basis’) patent C.



Seven families following the ETSI family definition. Note that these are ego-families (as seem from one specific patent).

- Data on disclosed patents provides valuable insights into which patents are potentially essential to a standard. This data clearly serves an important aspect within the operation in SDO itself, namely, to ensure that a standard only requires known patented inventions for which a FRAND commitment has been issued. It is important, however, to be well aware of the intrinsic limitations of such data if used for other purposes. Among other things, patents disclosed as being potentially essential (1) may not be owned (anymore) by the disclosing firm, (2) may not be factually essential (3) may not be granted, (4) may not be enforceable (valid, non-expired, renewal fees paid, etc.), (5) may greatly differ in technical merit, (6) may relate to functionalities not relevant for a certain product category (e.g., a mobile phone or an infrastructure product), and (7) may relate to optional features that might not be used in a given device conforming to the standard. Moreover, the patent families of these patents may differ substantially in terms of geographical coverage.

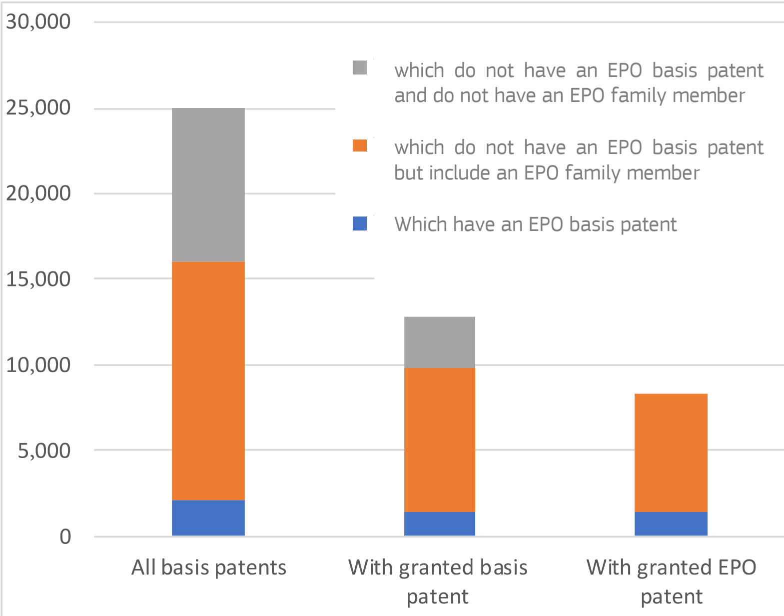
- ETSI-disclosed patents can be used as a starting point for an essentiality assessment procedure Overall, the dataset includes 25,070 disclosed patent families, of which 16,089 contain an EPO patent (Table 25 left column, second plus third row, also shown in Figure 29). Focusing on disclosed patent families whose basis patent was granted as of February 2019 yields a much lower total of 12,760 families, among them 9,818 with an EPO family member (Table 25, middle column). If one limits the assessment procedure to SEP exposure in Europe (i.e. only looks at patent families that comprise a granted EPO patent), then the relevant dataset includes 8,320 EPO patents that would need to be investigated: 1,449 families that have a granted EPO patent as a basis patent, and for 6,871 families there is a granted EPO patent in the family of the disclosed basis (Table 25, right column). It is worth noting that those 2,942 patent families whose basis patent is granted but that do not contain an EPO patent are mostly filed at the Chinese patent office and at the WIPO. Numbers could go down a bit if one excludes expired patents from the analysis, as well as patents that cannot be enforced because renewal fees have not been paid. But doing so requires careful consideration whether there are family members of that patent that are still alive, which would result in enforceability (and thus SEP exposure) in other geographies. Further note that these were the numbers as of February 2019[67] and, almost on a daily basis, new patents are disclosed, and patents that were disclosed earlier in time, are granted.

Table 25. Number of disclosed patent families (following the ETSI family definition): total, with granted basis patent and with granted EPO patent.

Patent families	Total number	With granted basis patent	With granted EPO patent
Total number of disclosed families (patent families following the ETSI family definition)	25,072 (100%)	12,760 (100%)	8,320 (100%)
which have an EPO basis patent	2,151 (8.6%)	1,449 (11.4%)	1,449 (17.4%)
which do not have an EPO basis patent but include an EPO family member	13,938 (55.6%)	8,369 (65.6%)	6,871 (82.6%)
which do not have an EPO basis patent and do not have an EPO family member	8,983 (35.8%)	2,942 (23.1%)	n/a

67 These numbers are based on the ETSI database as of late February 2019 (when we retrieved that data), and patent grant information is from the PATSTAT 2018 Autumn Edition.

Figure 29. Number of disclosed patent families (following the ETSI family definition): total, with granted basis patent and with granted EPO patent.

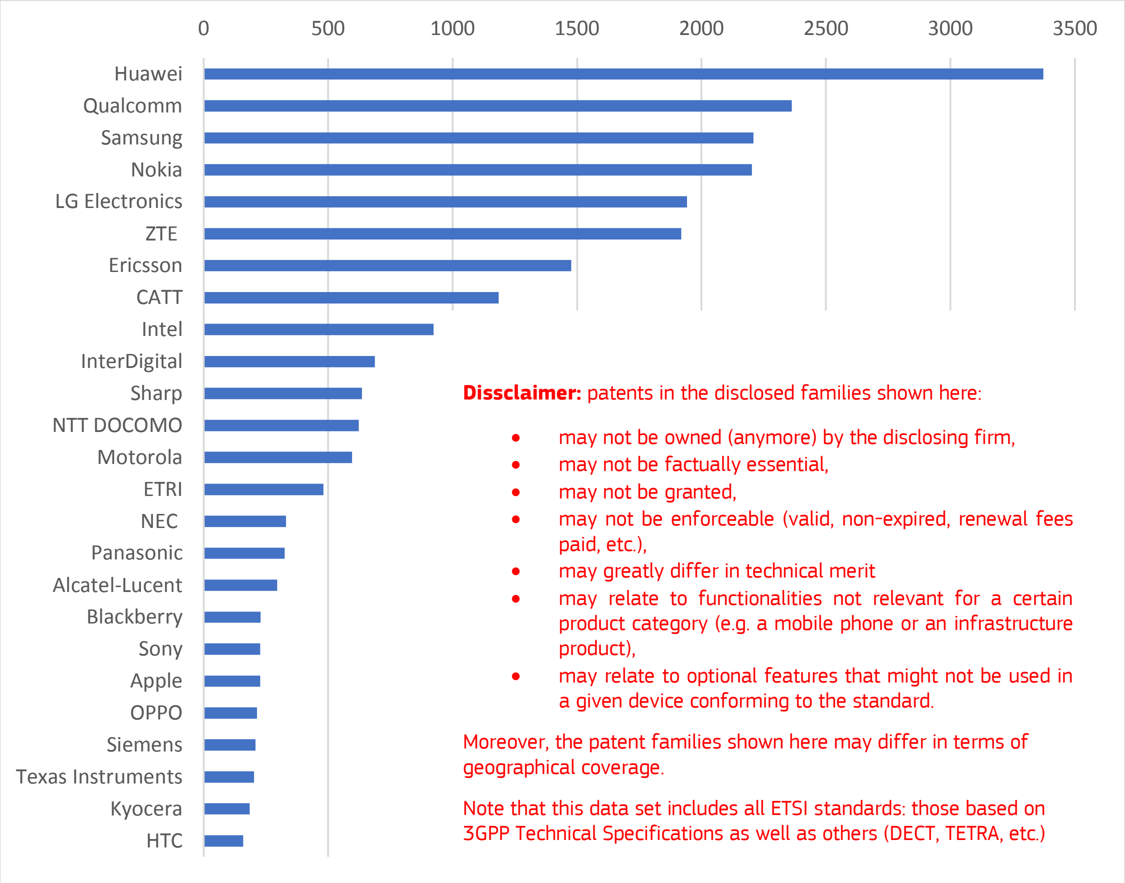


- For the 6,871 families where there is a granted EPO patent in the family of the (non-EPO) disclosed basis patent, it is key that the ETSI patent family definition is used to identify that EPO patent. If the DOCDB family definition were used instead, changes are significant that no EPO patent is identified at all while there may very well exist a factually, EPO family member of the disclosed patent. If the INPADOC family definition were used, changes are significant that an EPO patent is selected which is factually non-essential, while another EPO family member exists that is factually essential (bot not selected).

- There is considerable variety over time in the number of new patent families being disclosed. In general, the trend us upwards (see Figure 8, above, for more details). In the years 2017 and 2018 alone, 9,311 new patent families were disclosed to ETSI – that is 37% of the total of all 25,072 families (the year 2019 falls mostly outside our data set). This is a clear peak compared to other years, most likely because of intense 5G standardization activity during that period.

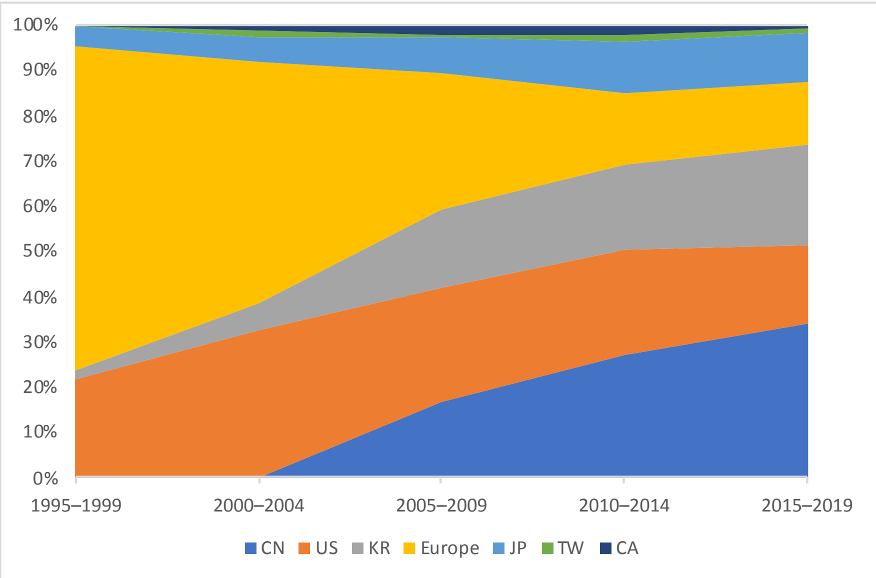
- There is considerable fragmentation in the distribution of companies – or organisations – that disclosed these patents, as shown in Figure 30. Furthermore, the distribution is skewed, with a long tail: out of the 148 disclosing firms, 14 disclosed more than 500 patent families, 21 companies disclosed between 50 and 500 families, and 113 disclosed less than 50 families. Again, we emphasize that the current owners of these patents may be different from the disclosing firms. These distributional characteristics are relevant in case any essentiality scheme chooses to use sampling: given the skewed nature of the distribution, a sample of disclosed patents might need to be stratified in order to be representative of individual firms. Assessing larger samples or even full portfolios for such companies is a potential solution.

Figure 30. Number of disclosed patent families (following the ETSI family definition) by earliest disclosing company; shown for the 35 largest companies; full time period (1990-2019). (data set: the 25,072 disclosed basis patents, coming from different jurisdictions)



- We observe big shifts over time in terms of the home country of firms disclosing patents (see Figure 31 and Section 3.5, above, for more details). Especially remarkable is the recent increase in shares of disclosed patent families from Chinese firms and, to a lesser degree, from South Korean firms, at the expense of European and US firms. Again, we stress that such numbers as such do not say yet anything about whether these patents will eventually be granted (and in which countries), what their technical merit is, etc. But the growth in numbers is notable. This is also reflected by an increasing share of disclosed basis patents from the Chinese patent office (see Section 3.6 for details).

Figure 31. Share of disclosed patent families (following the ETSI family definition) by home country/region of disclosing firm, over time (data set: the 25,072 disclosed basis patents, coming from different jurisdictions).



Note: the 1990-1996 time frame is omitted in this picture, as there are few observations in that time slot.

- It would be desirable to break ETSI disclosures related to cellular standards up into, for instance, technology generations (2G, 3G, 4G and 5G). Yet, despite the high sophistication of the ETSI disclosure database, for the bulk of the ETSI disclosures, such a distinction cannot be made in a reliable way without examining the content of each individual disclosed patent. This is related, among other things, to the way 3GPP technical specifications series are structured (see Section 2.7 for details).

- This study also investigated whether SDO disclosed patents differ in quality from comparable other patents. In this context, we looked at two dimensions of quality: technical merit (i.e., the technological contribution to the field) and economic value (i.e., the economic returns that a patent generates). We use observable characteristics of patents and patent families, such as forward citations, family size, and patent renewal, to proxy both dimensions. While such proxies are known to be far from perfect, they still do provide us with valuable insights. Our analyses find that disclosed SEPs score higher in terms of all the main proxies for patent quality that are commonly used in the patent literature. This is true for both those variables usually associated with technical merit and those variables usually associated with economic value. All in all, these results confirm the expectation that patents disclosed to ETSI are of greater technological importance and have more market potential than suitable controls.

Our interpretation of these findings is as follows:

- That disclosed patents have a *higher economic value* than comparable, non-disclosed patents is not surprising. After all, a subset of these patents will become factually essential patents, will then need to be licensed by all parties in the world that implement that standard,[68] and thus have the potential to generate significant licensing revenues (or other benefits such as cross-licensing opportunities). Hence, it is rational for applicants of such patents to seek protection in many countries and renew their patents, which are exactly the variables that are associated with higher economic value of a patent.
- That disclosed patents have a *higher technical merit* may indicate that SDOs are able to attract attractive technologies (to know whether they are able to *select* attractive technologies for the standard, we would need to know which of these are factually essential). Yet, the higher technical merit may also be the *result* of the act of disclosure as such. Firstly, disclosed patents are more 'visible'. Secondly, for any party investing in R&D in a technical field where standards are important, it is rational to build such R&D upon knowledge already embedded in these standards, instead of on a 'dead track'. This also increases the likelihood that essential patents (and also the wider set of disclosed, potentially essential patents) receive more citations from future patents.

[68] To be precise: insofar the product implementation is in a product category that requires the essential patent, and it is not only essential for an optional feature that is not implemented in the product implementation.

6 References

Bekkers, R., Catalini, C., Martinelli, A., Righi, C., and T. Simcoe (2017) "Disclosure Rules and Declared Essential Patents", NBER Working Paper No. 23627. Available at http://www.nber.org/papers/w23627.

Bekkers et al. (2020). Pilot study for essentiality assessment of Standard Essential Patents. Study commissioned by the European Commission EC Joint Research Centre (JRC) in collaboration with DG GROW and DG CONNECT. Available at https://ec.europa.eu/jrc.

Branstetter L. (2005), "Exploring the Link between Academic Science and Industrial Innovation", Annals of Economics and Statistics, 79/80: 119-142.

Cassiman, B., R. Veugelers, and P. Zuniga (2008). "In search of performance effects of (in)direct industry science links", Industrial and Corporate Change, 17(4): 611–646.

Criscuolo, P. and B. Verspagen (2008), "Does it Matter where Patent Citations Come from? Inventor vs. Examiner Citations in European Patents", Research Policy, Elsevier, No. 37 (10): 1892-1908.

ETSI (2018a). Rules of Procedure of the European Telecommunications Standards Institute Version approved by General Assembly#70 (SCM) (18 April 2018). Part of the ETSI DIRECTIVES, Version 39, 8 October 2018.

ETSI (2011). ETSI IPR ONLINE DATABASE: User Guide for Anonymous Users. Dated 03/03/2011, Retrieved 22 February 2019 from https://ipr.etsi.org (at "User Guide").

EPO (2018). Data Catalog PATSTAT Global, Version 5.12 (for PATSTAT 2018 Autumn Edition), dated 01 October 2018.

ETSI (2019). 3GPP Projects for Specs Versions, V5. Document ETSI IPR(19)031_004a1, prepared by the ETSI secretariat.

European Commission (2017). Communication from the Commission to the European Parliament, the Council and the European Economic And Social Committee: Setting out the EU approach to Standard Essential Patents' COM(2017)712 Final, 29 November 2017.

Hall, B.H., Jaffe A. and M. Trajtenberg (2005), "Market Value and Patent Citations", RAND Journal of Economics, No. 36, Spring.

Harhoff, D., F. M. Scherer, and K. Vopel (2003), "Citations, Family Size, Opposition and the Value of Patent Rights", Research Policy, 32(8): 1343-1363.

Henderson, R., Jaffe, A. and M. Trajtenberg (1998). "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." Review of Economics and Statistics, 80(1): 119-27.

Kuhn, Jeffrey M.; Younge, Kenneth A. & Marco, Alan C. (forthcoming). Patent Citations Reexamined. The RAND Journal of Economics.

Lanjouw J. O., A. Pakes, and J. Putnam (1998), "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data", Journal of Industrial Economics, 46(4): 405-432.

Lanjouw, J. O. and M. Schankerman (2001), "Characteristics of Patent Litigation: A Window on Competition", RAND Journal of Economics, 32(1): 129-151.

Lanjouw, J. and M. Schankerman (2004), "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators", The Economic Journal, 114(495): 441-465.

Lerner, J. (1994), "The Importance of Patent Scope: An Empirical Analysis", RAND Journal of Economics, 25(2): 319-333.

Narin, F., K.S. Hamilton, and D. Olivastro (1997), "The increasing linkage between U.S. technology and public science". Research Policy, 26: 317-330.

Mariani, M. (2004). What determines technological hits? Research Policy, 33(10), 1565–1582.

Pakes, A. and M. Schankerman (1984), "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources", in Griliches, Z. (ed) "R&D, Patents, and Productivity", University of Chicago Press.

Rysman, M. and T. Simcoe (2008), Patents and the Performance of Voluntary Standard-Setting Organizations, Management Science, 54, issue 11, p. 1920-1934.

Sherry, Edward F. and Teece, David (2004). Royalties, evolving patent rights, and the value of innovation. Research Policy, 33 (2), 179-191.

Squicciarini, M., H. Dernis and C. Criscuolo (2013), Measuring Patent Quality: Indicators of Technological and Economic Value, OECD Science, Technology and Industry Working Papers, 2013/03, OECD Publishing. http://dx.doi.org/10.1787/5k4522wkw1r8-en.

Trajtenberg, M. (1990), "A Penny for Your Quotes: Patent Citations and the Value of Innovation", RAND Journal of Economics, 21 (1): 172-187.

WIPO (2008). WIPO Intellectual Property Handbook. WIPO Publication No. 489 (E), Second Edition.

Appendix 1: Name harmonization for disclosing firms

For purposes of readability, we removed indications to legal forms such as "Inc."; "Corporation", "Ltd.", "GmbH", etc., from the firm names. In some cases, we added (for instance, 'IRT' was changed in 'IRT (Institut für Rundfunktechnik).[69]

Sometimes, companies have used different names when disclosing patents to ETSI (different spellings, legal entities from different countries, names of full subsidiaries). Where we believed this was the case, we brought these entries together. We also did so when a company changed name in its entirety (e.g. 'Research in Motion' who renamed itself into 'Blackberry'). **Table 26** shows when we merged two or more names of disclosing entities.

Note, however, that our starting point was the name as provided at the time of disclosure, and we did not bring entries together that may have been part of mergers or acquisitions (M&A's), that went bankrupt and assets transferred to new owners, or cases where patent ownership of parts of full portfolios was transferred, or otherwise. Such an exercise falls out of the scope of our study. (Names of merged entities may appear in our data, however, when firms themselves used these names when they submitted disclosures. For instance, our database has separate names for "AT&T", "Lucent", "Alcatel", "Alcatel-Lucent" and "Nokia", where all of these entities was part of M&A's.)

Table 26. Overview of merged company names

Short name	Original disclosure name
Alcatel-Lucent	'ALU', 'ALU', 'ALCATEL-LUCENT', 'ALU PARENT', 'ALU-ALU Shanghai Bell', 'Alcatel-Lucent Deutschland AG', 'ALU USA', 'Alcatel-Lucent Shanghai Bell', 'ALU US', 'ALU and ALU-ALU Shanghai Bell Co. Ltd'
Apple	'Apple Inc.', 'Apple (UK) Limited'
AT&T	'AT&T', 'AT&T Intellectual Property II, LLC'
Blackberry	'Research in Motion Limited' ,'BlackBerry UK Limited' ,'BlackBerry LTD
Dolby Laboratories	'Dolby Laboratories Inc.', 'Dolby Laboratories, Inc.'
Ericsson	'Ericsson', 'ERICSSON MOBILE COMMUNICATIONS'
Huawei	'Huawei', 'Huawei Technologies Co. Ltd.', 'Huawei Technologies Co., Ltd.'
INFINEON	'INFINEON TECHNOLOGIES', 'Infineon Technologies Flash'
Innovative Sonic	'Innovative Sonic Ltd.', 'Innovative Sonic Corp.'
InterDigital	'InterDigital Technology Corp.', 'InterDigital Patent Holdings', 'InterDigital Patent Holdings, Inc.', 'InterDigital Technology Corporation', 'IPR Licensing Inc.', 'IPR Licensing, Inc.', 'ITC', 'IDPA Holdings, Inc.'
KPN	'KPN N.V.', 'Koninklijke KPN N.V.'
Mitsubishi Electric	'Mitsubishi Electric RCE', 'Mitsubishi Electric Info', 'Mitsubishi Electric Corp', 'Mitsubishi Electric Telecom'
Motorola	'Motorola Mobility Inc.', 'MOTOROLA Inc', 'MML'
Nokia	'Nokia Corporation', 'Nokia Technologies Oy', 'Nokia Networks Oy', 'Nokia Networks', 'NOKIA MOBILE PHONES', 'Nokia Shanghai Bell', 'NOA'
Orange	'ORANGE', 'Orange Personal Communications'
Panasonic	'Panasonic Mobile Communication', 'Panasonic Corporation'
Qualcomm	'Qualcomm Incorporated', 'QUALCOMM Inc', 'Qualcomm Atheros, Inc.'
Blackberry	'Research in Motion Limited', 'BlackBerry UK Limited', 'BlackBerry LTD'
Siemens	'Siemens AG', 'Siemens Aktiengesellschaft'
Sony	'Sony United Kingdom Ltd', 'Sony Corporation'
Texas Instruments	'Texas Instruments Inc.', 'TEXAS Instruments'
Vodafone	'Vodafone IP Licensing Limited', 'VODAFONE LTD'

Note: sometimes multiple original disclosure names look almost identical, but there is a different comma, period or space in the name.

[69] This added information is primarily based on the online ISDL forms the firms submitted to ETSI, and the underlying facsimile documents, which often offer additional information on the name of the company.

Appendix 2: Companies by total number of disclosed patent families

Table 27. Companies by total number of disclosed patent families (following the ETSI family definition). **See also the disclaimers show in Figure 14.**

Disclosing firm	1990 - 1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Huawei							66	54	19	45	322	80	594	151	149	175	101	122	402	727	366	3373
Qualcomm			154		1	43	7	96	68	92	161	211	138	151	52	104	96	184	181	624		2363
Samsung					75			46	17	99	88	53	230	155	131	260	141	26	365	523		2209
Nokia	27	24	160	106	65	45	85	67	70	124	205	130	113	86	134	144	127	87	62	342		2203
LG Electronics	7							2			237	23	119	159	1	219	85	163	832	95		1942
ZTE												70	82		391		208	51	11	1103	3	1919
Ericsson	37		2	119		86	56	3	1		114	142	40	59	97	91	161	102	294	72		1476
CATT													258			94	19		130	685		1186
Intel													7	11	1	35	6	73	694	96		923
InterDigital			13			86			65	120	35	106	56	28	65	42	20	35	17			688
Sharp													7	55	49	50			347	128		636
NTT DOCOMO										49	71	76	48	22	70	44	27	56	122	38		623
Motorola	40	6	3	60	37	4	2	9		1	27	256		70	81							596
ETRI					6					1	30		26	1	17	13	46	8	321	12		481
NEC	3					41					11	1	34	93	15	19	31	29	37	17		331
Panasonic							40	12			33	46	5		81		85			24		326
Alcatel-Lucent									1	5	11	10	24	7	21	24	9	19	50	115		296
Blackberry				2				1	14	2	5	20	13	43	37	37	6	19	13	15	2	229
Sony											25	2				63	34	38	17	49		228
Apple									13				22	93	19	20	16	23	14	8		228
OPPO																			2	212		214
Siemens	2	1	25	8	22	4	11	28	23	7	13		7	4	2	14	13	9	13	2		208
Texas Instruments							6	1		30			43	71	29		22					202
Kyocera														10	67			108				185
HTC													24	41	23			27		44		159
Fujitsu							2											93	64			159
Innovative Sonic				17			12						80		13					20		142
Philips	16				45						21	1	7			39				12		141
KT Corp.																17		46		62		125
MediaTek																12		40		40		92
Nokia Siemens Networks (NSN)																			20	53		73
Fraunhofer IIS									2									58	1	7		68
Alcatel	25	2		23	6	1	3	1														61
Nortel Networks			1	1		7	5		2	31	1	5	7									60
Pantech														41	13	5						59

(continued on the next page)

Disclosing firm	1990 - 1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
ITRI		2														15			18	17		52
Orange		2		1		3	5		2	3		2	4	7	6	4	2					41
Microsoft					1										33							34
Mitsubishi Electric	6			2	14									12								34
Dolby Laboratories						24												9				33
KPN	1		2		2			1		2		1		3	2	4	4	5	1	3		31
Telia	29																					29
Toshiba						21										4	2					27
General Dynamics								7	3					16								26
Convida																			26			26
ASUSTeK				5				1												19		25
IDAC Holdings																			22			22
Unwired Planet															14	1	3		3			21
Hughes Network Systems					19																	19
Renesas														6	12							18
VoiceAge			2			1						1						14				18
Thomson																16			1			17
Marvell												7			9							16
Deutsche Telekom	1						1		5		3			1			4					15
Evolium				3			10	1														14
ITL (Innovative Technology Lab)																				13		13
INFINEON						3		1		9												13
Freescale Semiconductor												7	5									12
Sun Patent Trust																				12		12
Inmarsat				3													9					12
Sierra Wireless														1				1	4	5		11
Hanyang University																6		5				11
CP8 Technologies	10																					10
Optis Wireless Technology																5	4		1			10
OpenTV													10									10
AirTouch Communications	10																					10
SKT														6			3			1		10
Other 81 firms	19	16	6	11	13	2	9	10	12	16	7	5	9	2	11	10	7	20	21	10	0	216
Totals	233	53	368	361	306	371	320	341	317	636	1420	1255	2012	1405	1645	1586	1291	1470	4106	5205	371	25072

Notes: *This table shows all firms with 10 or more disclosed patent families (following the ETSI family definition). Data for other firms are aggregated and shown at the bottom of the table. For two of the 25,072 patent families, the earliest disclosure was dated "1900". Assuming this is an error, we changed that to 1990, the earliest date of other patents in the disclosure database.*

KJ-02-20-569-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/313626

ISBN 978-92-76-20786-3