Assessing the Influence of ERC-funded Research on Patented Inventions

Final report

11 November 2022

Prepared by a group of independent experts.

Group members:

Federico Munari*, Herica Morais Righi*, Maurizio Sobrero*, Laura Toschi* *(*Alma Mater Studiorum Università di Bologna)*

Ellsa Leonardelli**, Stefano Mainini**, Sara Tonelli** *(**Fondazione Bruno Kessler)*

Contents

Figures

Tables

1. Introduction

Context and background of the study. In the framework of the European Research Council (ERC) mission to reinforce the excellence, dynamism and creativity of European research, the ERC Monitoring and Evaluation Strategy has identified "impact beyond science" as one of its objectives, in line with the ERC mission. Socio-economic impact is at the core of the three dimensions of this objective: economic benefits, societal benefits, and policy making. One way to measure and assess the impact beyond the science of ERC-funding is to analyse the contribution of ERC-funded researchers to various measures of science valorisation activities (such as the creation of new startup companies, licensing agreements, research collaborations, public engagement initiatives). These are considered the typical direct channels through which the effects of publicly funded research are transferred to the economy and society at large.

In parallel with assessing direct links between the ERC-funded research in a project and the channels that transfer the knowledge from the project to various socio-economic actors, there is a challenge in tracing the contribution of ERC-funded knowledge to the development of key innovations. With this overall objective in mind, the ERC has identified a stepwise approach. The purpose of this study is to accomplish the first step of this approach. It consists of an analysis of how ERC-funded research findings influence or enable the development of a set of technologies, through the investigation of ERC-supported scholarly publications being cited in subsequent patented inventions.

The results of this work will be used by other experts in a second step of the broader programme. They will be object of a different study, to perform a patent landscape analysis for a set of technologies that form the core of a small number of selected innovations (existing or forthcoming). The final aim of the complete exercise is to perform a qualitative analysis of the influence of ERC-funded research on the development of the identified innovations.

Objectives of the study. The objective of this study is to identify and analyse the patents that can be linked to ERC-funded projects (through citations to the publications generated by such projects) and subsequently classify them in large technology areas. Following consolidated approaches in the economics of innovation literature, we worked on two different indicators of knowledge development: scientific publications and patents. The simple intuition is to measure if, how and to what extent the new discoveries presented by ERC grantees in scientific publications inspired new technologies described in new patents. We used the publications cited in the non-patent literature (NPL) section of patents applications to trace this influence. A complementary set of analyses were also performed for patents reported at the ERC Executive Agency by the Principal Investigators (PIs) as a direct outcome of the ERC project (that we label "self-reported" patents).

Similarly to scientific publications, also in patent documents references are provided in order to delimit prior art and prove the novelty of the invention in view of the existing technological development. These references mostly concern earlier patents, but they also include non-patent literature (NPL), such as scientific publications. Although citations to articles contained in patent documents are not perfect measures of knowledge flows, they have been increasingly adopted in the economic literature as signals of the intellectual influence exerted by public science on subsequent technological advancements. Therefore, they are a relevant instrument for capturing a broader effect and an additional reach of ERC-funded research, beyond the direct production of patents by ERC-funded scientists themselves[1].

In addition to quantifying the number of patents relying on ERC-funded research (both directly, as self-reported patents, and indirectly, by looking at patents citing ERC-funded publications), the report aims to better understand their distribution across time, across type of projects and type of applicants. Particular attention will be devoted to analysing the technological diversity of such patented inventions, in order to devise an initial map of the broad technological domains that rely on the scientific results of ERC-supported projects. To do so, we refer to the classification of the World Intellectual Property Office (WIPO), which is centred on 35 technological fields and related technological macro-sectors (5 sectors). This classification has the advantage of being established in existing patent databases, as well as highly adopted in both policy analyses (for instance, in the OECD statistics of patents by technology at the country level) and the existing literature on patent-to-paper citations. We also performed additional examinations in order to identify patents related to climate change mitigation and adaptation technologies, alongside patents related to digital transformation technologies, given that both the green transition and the digital transition represent strategic priorities for shaping not only the EU's political and industrial agenda, but the future of Europe in general.

The analyses consider the whole set of ERC-funded projects activated during the FP7 programme (2007-2013) and during the first three years of the H2020 programme (2014-2016), in order to have a sufficient time frame to capture patent citation dynamics. Due to the considerable time-lag between generating research outcomes and exploiting the innovation through patents, the expected coverage for both citing and self-reported patents remains still underreported for more recent projects (and in particular for H2020 projects), and it is expected to increase significantly in the coming years. The analyses cover projects over the entire

[1] Patents are only one way through which we can trace the impact of new scientific discoveries and can still be considered as intermediate output of the innovation journey. They present the advantage of being a measurable indicator linked by definition to technical inventions. In addition, in the course of the application and examination process, applicants are requested to disclose "prior art" against which the focal invention is distinguished and upon which the inventors may have relied on their own inventive process (Marx and Fuegi, 2020).

spectrum of scientific sectors supported by ERC funding: Life Sciences (LS); Physical Sciences and Engineering (PE); Social Sciences and Humanities (SH).

Methodology. The study methodology consisted of four different phases. In the first phase, we first identified ERC-funded grants for the period of analyses (4,556 ERC FP7 grants for the period 2007-2013; 2,115 ERC H2020 grants for the period 2014-2016), using information provided by the ERC Executive Agency. In the second step, we identified the publications produced by these research projects using information reported in the Cordis and Scopus databases, complemented by information provided by the ERC databases. The third step involved identifying the patent applications that build on such EU-funded research publications, using patent-to-publication citation information to reconstruct linkages between grants and citing patents. To this end, we used information reported in the NPL (Non-Patent Literature) field of patents using the Patstat database, provided by the European Patent Office.

In the fourth and final step, we performed a set of analyses related to the distribution of linked patents by technology domains in order to map the influence exerted in different technological fields. To this purpose, we first adopted the WIPO technology classification to link patent IPC classes to 5 macro-sectors and 35 related technology fields. We then performed additional examinations in order to identify patents related to climate change mitigation and adaptation technologies (relying on the EPO classification of such patents) and patents related to digital transformation technologies (adopting the EPO classification of 4th Industrial Revolution patents). Finally, we identified a set of ERC-supported projects whose publications were highly cited by subsequent patents, so as to illustrate a selected number of technological fields where the influence exerted by ERC research was particularly prominent. We replicated the same type of analyses for the set of patented inventions directly reported by beneficiaries to the ERC as an outcome of the project. Information on such self-reported patents were directly provided by the ERC Executive Agency, and then further complemented by additional information from the Patstat database.

Structure of the report. The report is organised as follows: Section 1 is the Introduction. Section 2 presents an executive summary of key findings. Section 3 describes the data sources and the methodology of the study. Section 4 details the main findings related to the general characteristics of citing patents and self-reported patents. Section 5 describes the key insights related to the analysis of citing and self-reported patents by technology domains. Section 6 concludes with a set of recommendations and ideas for future developments. The Annexes at the end of the document present further details the methods and data that underpin the study.

2. Executive summary of main findings

- The report is based on data retrieved from 6,671 ERC-funded projects from all scientific sectors (LS, PE, SSH), including both FP7 projects spanning the whole programme and H2020 projects that started between 2014 and 2016. We recovered 172,683 scientific publications linked to these projects and 34,513 patent applications citing those publications as references in the NPL.
- We also identified 1,550 self-reported patent applications, i.e., directly reported by beneficiaries at the ERCEA as an outcome of 700 of the 6,671 ERC projects. If we look at the number of projects with self-reported patent applications, they represent about 10% of the project population observed.
- The analyses show that ERC-funded projects generated considerable citation influence upon patentable technology. More than 40% of ERC grants generated research that was subsequently cited by patents. This is significantly higher than the share of grants directly generating self-reported patents (around 10%), offering us specific evidence of the indirect effect on technological development. Notably, this effect did not emerge from an analysis based solely on the specific outputs reported by each grantee.
- The large majority of patents citing publications from ERC-funded research in our dataset were linked to FP7 projects (32,728 patents linked to 2,370 FP7 projects, corresponding to a 52% share of FP7 projects cited by at least one patent application). This confirms that it takes time for science to influence technology and promote innovation. Considering the time lag between achieving research results and applying for patents, and assuming comparable trends to those observed for FP7 projects, the number of patent citations linked to H2020 projects is expected to increase significantly in the coming years.
- Coherently with the few other studies that have recently presented similar analyses in different countries and programmes, we find significant variation across scientific fields in the influence exerted by ERC-funded research upon technology development. The percentage of projects receiving patent citations via publications was higher for LS projects (61%), followed by PE projects (46%). As expected given the technical nature of the knowledge embedded in patented inventions, the share of projects from SSH sector being cited in patents was considerably lower (7%), although not null.
- On average, 7.45% of the publications produced by ERC projects are cited by subsequent patent applications. However, there is significant variation across scientific fields in the patent-to-paper citation percentages. The largest percentage of scientific publications cited in patents in our dataset, relative to the total number of scientific publications generated in that sector, was found in the LS sector (around 12%), followed by the PE sector (around 6.58%). Such numbers fall in the high-end of the range of average values reported by academic studies that have investigated scientific papers cited in patents (which, depending on the study, vary from 1% to 11%). That said, it is very difficult to make direct comparisons with existing

literature because the results are sensitive to several study-specific conditions (and, in particular, to the scientific field(s) under analysis).

- Although the majority of citations to ERC-funded publications (around 50%) come from patents assigned to firms, a significant share of citing patents are still owned (or co-owned) by universities and research organisations. The ownership patterns of self-reported patents show that universities (48% of cases) and research institutes (23% of cases) play a dominant role as applicants, with firms being the owner or co-owner of self-reported patents in 15% of cases.
- Looking at the distribution of citing patents in the 35 technology fields identified by the WIPO, we noticed a concentration of patents relying on ERC-supported science in a selected number of technology fields (biotechnology; pharmaceuticals; computer technology; organic fine chemistry; measurement; semiconductors). Together, these accounted for 68.2% of the total received patent citations.
- In line with the findings of more recent research on patent-to-paper citations, we found that patent citations to ERC-based research outputs are more common in technologies closer to the science frontier and in areas where industry has a greater science-based R&D orientation.
- We found generalised evidence that patent citations to ERC-funded research often flow across technological fields, coherently with the idea of widespread diffusion of frontier research results. This flow across technology domains was more widespread for projects funded in the Engineering and Physical Sciences sectors, while Life Science ones tended to be more concentrated.
- Focusing on the domains of climate change mitigation and adaption technologies, we document a prevalence of occurrences in two macro-fields: the sub-field “Reduction of greenhouse gas (GHG) emissions, related to energy generation, transmission or distribution” and the sub-field “Technologies for Adaptation to Climate Change”.
- We observed a focus on technologies that are relevant for the Digital Transformation, using the classification of Fourth Industrial Revolution (4IR) patents developed by the EPO. This suggests a significant specialisation of ERC-linked patents in this area, especially in the fields related to “Data Management”, “Smart Health” and “Connectivity”. As expected, the concentration of 4IR patents was particularly pronounced in the group of patents linked to PE projects. It is noteworthy that the largest share of patent applications citing research results from SSH projects was related to the 4IR technological domain.
- The qualitative analyses of ERC FP7 and H2020 projects whose publications were highly cited in patent documents illuminated some areas where the knowledge stemming from ERC projects was particularly valuable for inspiring subsequent technological developments. Such area included image recognition technologies; graphene applications; solar cell technologies; applications of microRNAs; immunotherapy treatments, and stem cell technologies.

- Our evidence supports previous findings regarding a positive relationship between scientific impact (as measured by the number of citations received in peer-reviewed publications) and technological impact (as captured by patent-to-paper citations).
- Our study confirms that the grants-publications-patents approach is a promising way to investigate the technological influence of publicly funded work. However, there are still areas where it could be improved, especially related to how citations are used by patent owners.
- In the era of Big Data, the quality of the information collected and its reporting require great attention and care. In this respect there would be value in the streamlining the formatting used by beneficiaries in order to improve the quality of information on patents self-reported by ERC grantees.
- Patents are only one way through which we can trace the impact of new scientific discoveries. After all, patents represent an intermediate output in the innovation journey. Further studies could integrate the grant-publication-patent flow with data on other direct measures of science valorisation.

3. Patent citations to scientific literature: an introduction

Patent documents, like scientific publications, included references to delimit prior art and prove the novelty of their invention in relation to existing technological developments. These references mostly concern earlier patents, but they also include non-patent literature (NPL), such as scientific publications. Prior art disclosure concerning scientific publications can be inserted by both inventors/applicants when drafting the application and by examiners in the course of the patent examination process. The requirements in this sense vary by jurisdiction.[2]

Exploiting this unique source of information, a vast body of studies in the field of economics of science and innovation have investigated patent citations to scientific literature as a way to better understand the complex relationship between scientific advancements and technology development. Pioneering work by Narin and colleagues (1997) and Jaffe et al. (1993) analysed patent-to-paper citations and sparked a wave of empirical studies examining the scientific intensity of patented inventions and the dynamics of knowledge flows from science to technology. One recent and promising segment of this research stream traces the reliance of patents on public funding sources, exploiting information from patent citations to publicly-supported scientific publications (Azoulay et al., 2019; Fleming et al., 2019; Li et al., 2018; Marx and Fuegi, 2020).

The purpose of this section is not to provide a detailed review of this literature (there are excellent review contributions in previous articles, such as Marx and Fuegi, 2020; Van Raan, 2017; Veugelers and Wang, 2019). Instead, we seek to summarise more recent contributions linking grants, publications and citing patents in order to better gauge the potential of the approach, provide some reference values to better interpret our own findings regarding ERC grants, and highlight the intrinsic limitations of this method.

The initial set of studies analysed the link between patents and publications using the NPL information to estimate the importance of science for technological innovation (Callaert et al., 2014; Jaffe et al., 1993; Tijssen et al., 2000; Veugelers and Wang, 2019). There has been a large volume of empirical studies taking patents as the starting point and comparing the characteristics and impact of those including science citations to those without (Cassiman et al., 2008; Fleming and Sorenson, 2004). Relatively less developed is the literature using scientific publications as the starting point, examining what determines their likelihood of being referenced in patents. Looking at more recent studies in this area, there has been some variation in the extent of science-to-patent linkages documented, which reflects underlying differences in study design,

[2] In the US patent system, for instance, there are strict requirements for both applicants and examiners to submit any prior art known to them at the time of application or examination, whereas in other patent jurisdictions, such duties are less pronounced.

context of analyses, data sources, and matching techniques (see Figure 1 for a summary of a selected number of recent studies).

Figure 1 - Selected recent studies analysing patent citations to scientific publications

Article	Sectors	Scientific Publication Data	Patent Data	Main Results
Ahmadpoor and Jones (2017)	Science and Engineering Sectors	Web of Science (WoS) publications	US granted patents	10% of scientific publications are directly cited by patents. The percentage increases significantly considering indirect connectivity. Significant variation across fields.
Veugelers and Wang (2019)	Natural sciences and engineering	Web of Science (WoS) publications in 2001	All patents in Patstat	On average, 11% of sample publications are cited in subsequent patents. Publications that score high in novelty are more likely to be cited in patents.
De Moya-Anegon et al. (2020)	All scientific sectors	Scopus publications (2008-2017)	All patents in Patstat	On average, 3.2% percentage of papers cited in patents. Large differences observed across disciplines.

Some studies have found that being cited by a patent is quite a rare event for scientific publications. For instance, in their study on the discovery of introns, Winnik et al. (2013) found that only 1% of intron-related WoS publications in the period 1986–2001 were cited in intron-related patents. In a recent review of this literature, Van Raan (2019) stated that only a small minority (around 3-4%) of publications covered by the Web of Science or Scopus are cited by patents, but this percentage is considerably higher (15%) for publications based on university-industry collaborations. De Moya-Anegon et al. (2020) arrived at a similar value when analysing a broad set of scientific sectors, including the social sciences and humanities. Drawing from Scopus (publications) and Patstat (for patents) data for the time period 2008-2017, their study reported that 3.2% of papers were cited in patents.

Other recent studies focusing on the science and engineering sectors have documented slightly higher overall average values. Veugelers and Wang's (2019) study—based on WoS journal articles from natural sciences and engineering published in 2001, alongside all the patents in Patsat—found that 11% (on average) of the sampled publications were cited in subsequent patents. The study by Ahmadpoor and Jones (2017), based on 4.8 million US patents and 32 million WoS research articles in science and engineering, found that around 10% of scientific publications were directly cited by patents, although such a percentage becomes definitively higher when considering multiple scientific links (i.e., the original publication is cited by another publication, which is then cited by a patent). Importantly, their study documented dramatic variation by technological fields in the patent propensity to cite science, with the technological fields closest to patent frontier including combinatorial chemistry, molecular biology, superconducting technology and artificial intelligence[3]. Similar evidence has been provided by Jefferson et al. (2018), who linked 11.8 scholarly outputs extracted over the 1980 to 2015 period to patent data included in Lens, showing an average of around 10% publications cited in patents.

There are some important indications emerging from such studies, related to complex interactions between science and technology that can be captured by looking at patent-publication citations. First, the number of citations to scientific literature in patents varies dramatically across technology fields; it is much more pronounced in fields at the science-to-technology frontier and in emerging fields. Second, the time dimension is important, as the time lag between the publication of scientific articles and their subsequent citation in patents may be significant and may differ substantially between the various fields of technology. Third, there are differences across patent offices (USPTO requiring relatively more NPL citations than EPO, due to stricter disclosure requirements). Fourth, the number of scientific references depends on several different factors (type of scientific and technological fields; type of applicants and inventions; time period under analysis; characteristics of the patent office; role of inventors versus examiners; matching approach). Thus, the average values of previous analyses cannot be used as actual metrics for comparisons, but rather as reference points to better interpret the results.

Taking these findings into account, we now turn to the second, and more recent, step of this approach: looking at the impact of public research funding on the production of new patented technologies (see Figure 2 for a summary of selected articles). This approach links specific research grants to subsequent patents using their publications. A seminal paper in this area is that of Azoulay et al. (2015), who developed a method to link NIH research funding to USPTO patents by pharmaceutical and biotechnology firms. The authors proposed a new methodology relating NIH grants to patents. Using data from NIH grants, they identified scientific publications by their acknowledgment data and linked those publications to patents through USPTO bibliometric references. In a subsequent related paper, Li et al. (2017) found that about 10% of NIH grants generate a patent directly, but 30% generate articles that are subsequently cited by patents.

[3] In this study, the scientific fields closest to the patent frontier include nanotechnology, materials science and biomaterials, and computer science hardware and software.

Other recent studies have documented the importance of public research funding on the generation of subsequent private-sector patents in the health sector. These studies show that public research investments have a positive effect on private-sector patent activities (Du et al., 2019; Jefferson et al., 2018; Poege et al., 2020). For instance, in one of the first studies to apply this approach outside the U.S., , Jefferson et al. (2018) applied the methodology to the U.K.'s Medical Research Council (MRC)-funded scholarship. The authors uncovered that subsequent patents cited around 10% of the publications supported by this institution (this study did not report information at the grant level, however).

Figure 2 - Selected recent studies analysing patent citations to scientific publications linked to publicly funded grants

Article	Institution Investigated	Sector	Scientific Publication Data	Patent Data	Main Results
Li et al. (2017)	U.S. National Institutes of Health grants (NIH)	Biomedical research	Publications listed in PubMed	US granted patents	10% of NIH grants connected to a patent directly, 30% of NIH grants generate articles cited by patents.
Jefferson et al. (2018)	U.S. National Institutes of Health grants (NIH)	Medical Research	PubMed from 2008 onwards	Global patent applications	Around 9% of publications linked to MRD funding have been cited by patents. No reference for patent citations at the grant level.
Fleming et al. (2019)	U.S. Federally supported research	Not specified	Web of Science (WoS)	US granted patents	13.75% of the patents generated from 1926 and 2017 are connected to a Federal research funding.

The practice of connecting research grants–scientific articles–patents is thus a fast-growing and very promising area of interest in the literature. That said, it is important to highlight the approach’s limitations and areas of improvement in order to better understand its value in mapping science-technology linkages. In this sense, we want to emphasise some lessons emerging from research in this area:

- There is very high variation in the percentage of patents citing papers in NPL across scientific and technological fields (Ahmadpoor and Jones, 2017; Marx and Fuegi, 2020).
- The process of translating new scientific knowledge into new patented inventions is long-term, complex and cumulative; thus, the majority of connections between patent and papers are indirect (Ahmadpoor and Jones, 2017).

- There might be unsystematic declaration of funding information in publications, thus limiting the reliability of article-funding source matches; this issue extends to major publications' datasets (Grassano et al., 2017; Rai and Sampat, 2012).
- Scientific sources of inspiration are not always reported by inventors in patents (Callaert et al., 2014), and not necessarily in the NPL section but rather in the text of the patent (Bryan et al., 2019).
- Patent citations to previous articles in patents might be added by patent examiners in the course of the patent examination process, rather than by inventors. In that case, the citations may play more of a legal role rather than reflect intellectual inspiration.
- There are significant differences between patent systems regarding the use of NPL in patents (Callaert et al., 2014).

Accounting for these limitations, the body of scholarly evidence highlights that patent citations to scientific literature can serve as valuable signals of the intellectual influence of public science on subsequent technological developments. This approach can be combined with previously used measures to reach a broader understanding of impact.

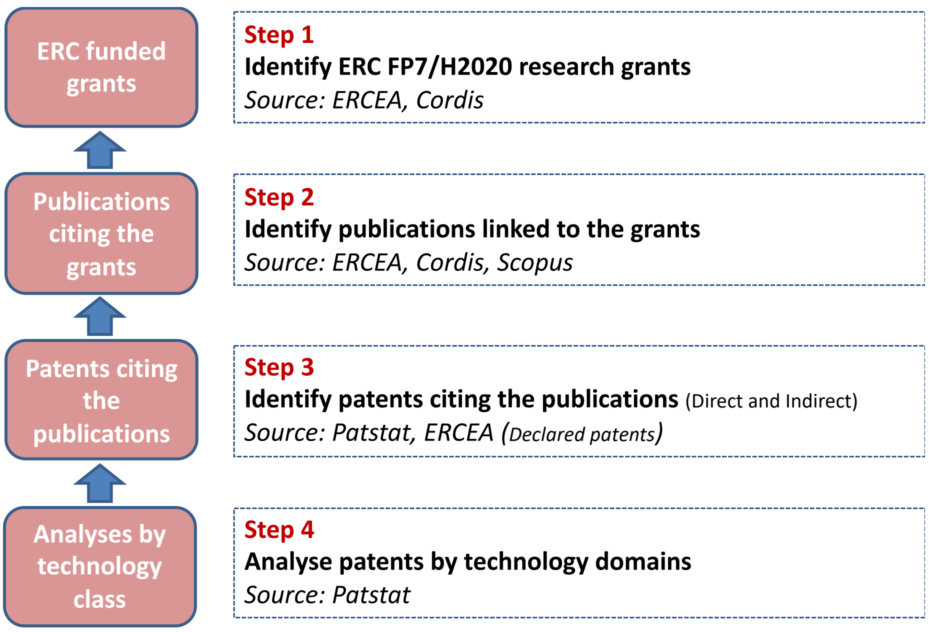
4. Methods and data collection

4.1 Overview of the study approach

The aim of this study is to analyse how ERC-funded research findings influence or enable technology development, through the assessment of scholarly publications being cited in patents. Here, we focus on the so-called Non-Patent Literature (NPL) field from Patstat, which allows one to identify scientific publications cited on the patent's front page or during the examination process. The objective is to identify and analyse the patents that can be linked to ERC-funded projects and subsequently classify them in large technology areas. We also performed a complementary set of analyses for patents directly reported at the ERCEA by the principal Investigators (PIs) of the ERC project ("self-reported" patents).

In order to complete the first step and build our database, we follow the innovation literature in crafting an approach to link ERC-funded research projects' contribution to the development of patents by tracing the patent citations to scientific publications. In order to identify patents linked to ERC research grants, we followed four steps, as shown in Figure 3.

Figure 3 – The approach of the study



We first identified ERC-funded grants for the period 2008-2022, using information provided by the ERCEA. We considered both projects funded by the ERC FP7 programme (for the period 2007-2013) and the ERC H2020 programme (available from 2014 to 2020). The second step involved identifying the publications produced by these research projects using the combination of Cordis[4] and Scopus[5] information, complemented by information provided by the ERC databases. We merged the information of such datasets using the DOI and/or title of each publication. Our third step involved identifying the EPO patents that build on such EU-funded research publications, using patent-to-publication citation information (using information about the NPL field of patents, as present in Patstat). Finally, we performed a set of analyses related to the distribution of linked patents by technology domains, in order to map the influence exerted in different technological fields. To this purpose, we first adopted the WIPO technology classification linking IPC patent classes to 5 macro-sectors and 35 related technology fields. We then performed additional examinations in order to identify patents related to climate change mitigation and adaptation technologies, as well as patents related to digital transformation technologies. We selected these two focal areas because both the green transition and the digital transition represent strategic priorities for the EU's political and industrial agenda. We replicated the same type of analyses for the set of patented inventions directly reported by the PIs to the ERC.

In the next section, we provide further details about the methodological approach used to construct the database.

4.2 Sources, data collection and methods

Identifying ERC-Projects

As a first step, we collected the list of projects supported by an FP7 or H2020 ERC-grant. A comprehensive list was provided by ERCEA. A total of 12,388 projects were identified: 4,556 FP7 and 7,832 H2020. For each project, we collected the following information: project-ID, acronym, title, call, grant type[6], scientific domain[7], start and end date, PI name, PI surname, Coordinator (Host Institution), and other participant institutions.

[4] Cordis is the Community Research and Development Information Service and it is the European Comissions' primary source of results from the projects funded by the EU's framework programmes for research and innovation. Information available at https://cordis.europa.eu/. We also integrated the data available on the Cordis website with data available on the European Research Council website.
[5] Elsevier's database of publications.
[6] Possible types of grants are: Starting Grants (StG), Consolidator Grants (CoG), Advanced Grants (AdG), Synergy Grants (SyG) and Proof of Concept (PoC).
[7] The possible scientific domains are: Physical Engineering (PE), Life Science (LS) and Social Sciences and Humanities (SH).

Creating a dataset of scientific publications supported by ERC projects

We next sought to create a dataset containing all scientific publications that have been supported by one (or more) of the ERC projects identified in the previous section, with their relative metadata (particularly relevant for use are *title*, *doi*, *authors* and *year of publication).* To this end, we retrieved a list of ERC-funded publications from five sources and organised them into a unique dataset: Cordis, Scopus, a dataset of publications provided by ERCEA and the publications collected in our previous research project "University research funding, patenting and technological impact", funded by the European Patent Office[8].

In particular, on the Cordis website, each ERC project has a dedicated page that lists the publications that have been reported directly by the PIs. The data from ERCEA and the previous EPO ARP research project are structured in a list of publications linked to their respective ERC projects. Scopus analyses the funding sources cited in the scientific publications' acknowledgments and inserts them into a dedicated metadata field. From Scopus we downloaded, via API, all the publications citing in their acknowledgments funding obtained from ERC FP7 or ERC H2020 projects (using also to this purpose the ERC project-IDS identified in the first step). Notably, the information available for each publication retrieved from Scopus was highly homogeneous and detailed. In case of multiple projects acknowledged in a publication, we linked that publication to each project.

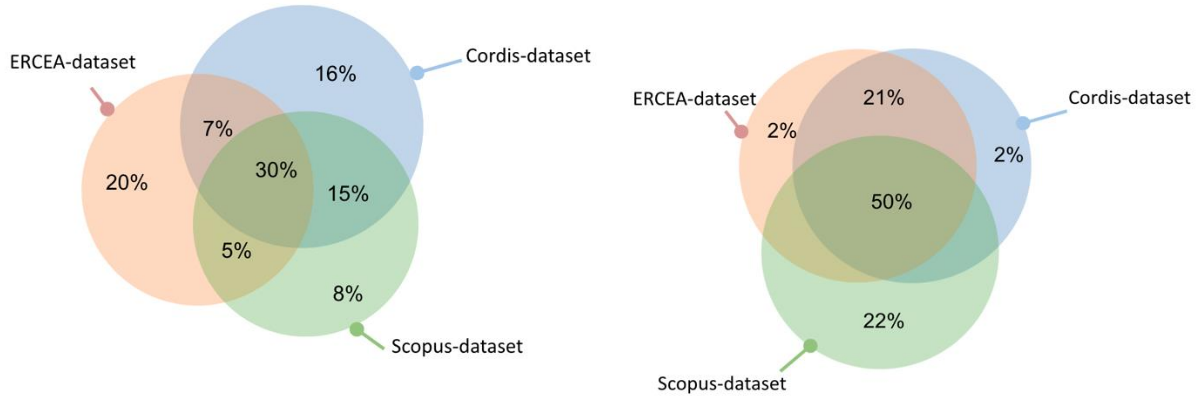
Collecting data from different sources allowed us to retrieve a larger number of publications, although this led to a partial overlap between the publications. To organise the five datasets into a unique dataset, we merged the overlapping publications from the various datasets, while maintaining the relative metadata from all the different sources. In order to identify the duplicates and merge them into a single entry, we matched the publications from each source based on the title and DOI. Publications with the same title and DOI, or just the same DOI, were considered to be the same and merged. To minimise false matches, titles were normalised to account for small differences in accents or quotes, while DOIs were checked to be properly formatted and accurate DOI codes. The Annex reports more details about the process and number of publications retrieved for each single dataset. After merging articles from the five data sources, the FP7 dataset counted 198,751 entries and H2020 dataset counted 97,875 entries.

Aside from retrieving more data, unifying the different datasets allowed us to measure the reliability of the entries. Indeed, publications found in more sources are more likely to be correct. From the merging of the different datasets, we observed the following: the publications from our EPO ARP previous research are

[8] This research project was funded by the Academic Research Programme of the European Patent Office, over the period 2018-2021. It was coordinated by the Polytechnic of Turin in collaboration with the Department of Management of the University of Bologna. We initially considered the publications databases from the portal for European data to double check our data collection process, but we ultimately did not use that in a systematic way for the analyses.

substantially all contained in the other datasets. Including this dataset was a sanity check control that confirmed that the current dataset is at least as good as the one from our previous research project. Figure 4 shows the overlap between the three principal sources of publications. We can observe how for H2020, there is substantial overlap between the Cordis and ERCEA datasets, while Scopus retrieved a big portion of publications independently. With the H2020 programme adding a mandatory acknowledgment section to each publication, retrieving publications (via acknowledgment section) from Scopus became more effective. It is possible that Scopus is able to retrieve publications that have not (yet) been reported by PIs to Cordis. Conversely, FP7 presents less overlap between datasets: FP7 projects are more distant in time, possibly making the retrieval of publications sparser.

Figure 4 - Overlap between three of the principal sources of publications utilised (left FP7, right H2020)



Although considering so many sources can introduce some noise, our ultimate goal was to maximise the probability of finding a publication in the Non-Patent-Literature of Patstat patents database (see next section). Here, we sought to retrieve and retain all possible information; during the next phase, we will implement a stricter check during the phase of matching these scientific publications with the non-patent-literature.

Linking the dataset of ERC publications to patents

After defining a dataset of relevant publications supported by ERC-grants, we proceeded to identify the patents citing those publications. To this end, we relied on the Patstat database to link publications and patents. Patstat is a database released by the European Patent Office (EPO) that provides access to patent data from more than 40 patent authorities worldwide. This database is updated every six months. At the start of the current project, the last available version of the database (and the one we used) was the 2021-Autumn edition.

We considered a link between a patent and an ERC publication when said publication is cited in the Non-Patent Literature section of the patents. Patstat provides a specific table that reports, for each patent, the citations to scientific publications included in the "Non-Patent-Literature" (NPL). To match the entries in the NPL table of Patstat with the scientific publications identified in the previous step, we considered three criteria: having the same first author's surname, the same title, and the same DOI. For NPL instances with structured metadata, we matched ERC publications' information only with the related Patstat field. In this step, we only considered publications resulting from exact matches (i.e., when the entire surname, title, or DOI was the same). When dealing with unstructured metadata, we compared the author, title, and DOI with all the information available in the metadata. While matches that occurred with two or more criteria (e.g., same title, DOI or same title and same authors surname, etc.) were considered reliable, matches that occurred only by DOI or title were manually controlled, while matches that occurred only because of the author's surname were discarded. The Annex provides more details on the methodology. A total of 36,901 matches for FP7 were found, involving 13,218 publications from our dataset and a total of 57,256 citations in the Patstat (the same ERC publication can be found more than once in the NPL table, but also the same entry in the NPL table can be cited more than once). For more recent H2020 projects, 3,521 matches were found. In particular, 2,127 different publications from our dataset of publications generated a total of 5,296 citations. The difference in matches between FP7 and H2020 is due to the more recent nature of H2020 projects (many of them were still ongoing or recently activated at the beginning of the study), which left a shorter time period for generating publications and being subsequently cited in patents.

After relating the NPL-table entries of Patstat to ERC grants, we proceeded to identify the linked patents' applications. We queried the Patstat database for each NPL publication found in the previous match to identify the linked patent's application number and additional information about the patent (as described in the Annex). We found a total of 40,346 patent applications citing a publication linked to ERC projects, where 37,885 were linked to ERC FP7 projects and 4,357 to H2020 ERC projects. Note that the total number of patent applications citing a publication from our entire dataset is lower than the sum of patent applications citing a publication from H2020 and FP7. This is because some patents cite publications from both programmes, thus creating an overlap.

Considered the information found on our sources, our final dataset exclusively included publications whose bibliographic information was available in SCOPUS, that started after the beginning of the project, and whose related patent applications were filed after the start of the project[9]. Moreover, as initially explained, we

[9] We decided to maintain only scientific publications for which information was available in Scopus (independently from the initial source of origin of such publication, being Cordis, the ERCEA and/or Scopus itself) so to have full bibliographic information on the publications (particularly the publication year of the article). We also decided to maintain publications that were originally identified only in the Scopus database (and that were not available in Cordis or in the ERCEA records) in those cases where we were able to identify the PI among the authors of such publications. This is a

maintained in the final dataset only H2020 projects (and the related publications and patents) that started in the years 2014, 2015, 2016, so there would have been a sufficient time span for generating publications and subsequent patents (in this respect, H2020 projects started after 2016 were probably still ongoing at the time of the study). Table 1 presents the main records from our final dataset used in the analyses reported in Sections 5 and 6. This final dataset includes 6,671 ERC projects (including 4556 FP7 projects and 2115 H2020 projects), 172,683 publications linked to them and 34,513 patent applications citing such publications.

Table 1 - Total number of ERC projects, related publications with Scopus Information, patents citing such publications, included in the final sample (all projects; FP7 projects; H2020 projects starting in the period 2014-2016)

Variable	Total Dataset	FP7	H2020 (2014-2016)
ERC Projects	6,671	4,556	2,115
Linked Publications*	172,683	134,961	40,407
Patent applications citing linked publications**	34,513	32,728	3,204

*Note: * A publication can be linked to more than one ERC-project (e.g. to a FP7 and to a H2020 project). For such reason, in this Table the sum of values related to FP7 and H2020 projects is higher than the total number of unique publications (reported in the column "Total Dataset").*
***A patent can cite publications linked to more than one ERC-funded project (e.g. to a FP7 and to a H2020 project). For such reason, in this Table the sum of values related to FP7 and H2020 projects is higher than the total number of unique patent applications (reported in the column "Total Dataset").*

Identifying self-reported patent applications

As a way to illustrate the main outcomes achieved, we also complemented the results obtained from the method based on patents-publications-grants matches with information on the set of patents directly reported at the ERC by the project's principal investigators. Such information is contained in the intermediate and final technical reports submitted to the Agency. The idea is to better understand how these approaches overlap or complement each other. To this end, we were able to use information directly provided by the ERCEA on the patents declared (or not) by each project. This internal database provided by the ERC contained 2,206 intellectual property records, 1,650 declared from FP7 projects, and 556 from H2020 projects[10]. From that list, we identified the patent application number and matched them to Patstat to gather information about application year, legal status, inventors, applicants, DOCDB family, IPC, NPL citations. From this search, we were able to identify 1,963 unique patent applications connected to either H2020 and FP7, related to 901 different projects. From these records, 1,572 patents were from FP7 projects (generated by 609 projects), while 446 patent applications were linked to H2020 projects (generated by 292 projects). However, we

conservative approach, given that Scopus does not always clarify the source of information used for identifying the funding sources behind a publication.

[10] As mentioned before, our dataset comprises ERC H2020 projects started between 2014 and 2016.

limited our final dataset around those filed after the start of the project. Table 2 presents the final dataset of self-reported patent applications that we used in the analyses, considering the period of time under analysis in this study (for the H2020 programme, we considered only self-reported patents from projects starting in years 2014, 205 and 2016).

Table 2 - Total number of ERC projects and related self-reported patents included in the final sample (all projects; FP7 projects; H2020 projects starting in the period 2014-2016)

Variable	Total Dataset (2008-2016)	FP7 Projects	H2020 Projects (2014-2016)
ERC Projects	6,671	4,556	2,115
Self-reported Patent applications*	1,550	1,340	222
Self-reported granted patents*	507	459	51

*Note: * A patent can be reported by more than one ERC-project. For such reason, in this Table the sum of values related to FP7 and H2020 projects is higher than the total number of unique patent applications/granted patents (reported in the column "Total Dataset").*

5. The characteristics of grants, scientific publications, and citing/declared patents

In this section, we describe the relevant characteristics of ERC grants (section 5.1), their scientific publications (section 5.2) and the linked citing and self-declared patents (section 5.3) included in our final sample, as explained in the previous section. As initially defined, this set of analyses focuses on all ERC FP7 grants; for ERC H2020 grants, we only considered those with starting years over the period 2014-2016, so to have a sufficient time span to capture the subsequent patent-publication citation patterns. Our data refer to 6,671 ERC projects linked to 172,683 publications and to 34,513 patents citing such publications in the NPL. If we consider self-reported patents (instead of cited patents), our 6,671 ERC projects reported a total of 1,550 patent applications. This section first briefly describes the characteristics of the FP7 and H2020 ERC-funded projects, which represent the objects of analysis. It then presents the publications generated from such projects, as described in the methodological section. The core of the section focuses first on the analyses of patents citing such publications as relevant background knowledge, then on patents directly reported at the ERCEA by the PIs as direct technical outcomes stemming from the projects.

5.1 Distribution of FP7 and H2020 ERC Grants

Among the 6,671 ERC projects, 4,556 were FP7 research projects from Life Science (LS), Physical Engineering (PE), and Social Science and Humanities (SH) sectors, while 2,115 were H2020 research projects for the period 2014-2016. The whole dataset contained 46.13% of projects from the Physical Sciences and Engineering domain (PE), 35.56% from the Life Sciences (LS) sector, and 18.03% from the Social Sciences and Humanities (SH) domain. The sample also included a limited number of projects (0.28%) corresponding to Synergy Grants, which are multidisciplinary in nature. If we look at FP7 and H2020 separately, the majority of projects in both cases derive from the PE sector (45.59% and 47.28% for FP7 and H2020, respectively), followed by LS (35.84% for FP7 and 34.94% for H2020) and SSH sector (18.17% of projects for FP7; 17.73% for H2020) (Table 3).

Table 3 – Total number of ERC projects in the sample, by ERC domain

Domain	N Total Dataset	% Total Dataset	N FP7	% FP7	N H2020	% H2020
Physical Engineering	3,077	46.13%	2,077	45.59%	1,000	47.28%
Life Sciences	2,372	35.56%	1,633	35.84%	7,39	34.94%
Social Science and Humanities	1,203	18.03%	828	18.17%	375	17.73%
Synergy Grant*	19	0.28%	18	0.40%	1	0.05%
Total	6,671	100.00%	4,556	100.00%	2,115	100.00%

* Synergy Grant projects are usually multidisciplinary, spanning diverse scientific domains, so we report them as a separate category. However, for 5 Synergy grants we were able to identify a unique scientific domain of reference. These 5 projects are thus reported in this table and in the following ones in the specific scientific domain of reference.

Parsing ERC grants by type of project, Starting Grants (StG) represented a share of 46.04% of all projects in the sample, followed by Advanced Grants (AdG) with a share of 32.11%, and Consolidator Grants (CoG) with a 14.38% share. Proof of Concept Grants (POC) accounted for 7.12% of the sample[11] and Synergy Grants (SyG) were 0.36%. The percentages of StG and AdG projects were higher for FP7 with respect to H2020 (51.19% vs. 34.94% and 37.51% vs. 20.47%), while CoGs and POCs were more represented in the pool of H2020 grants than FP7 (14.04% vs. 3.91% and 0% vs. 0.53%) (Table 4).

Table 4 - Total number of ERC projects in the sample, by ERC project type

Type	N Total Dataset	% Total Dataset	N FP7	% FP7	N H2020	% H2020
Starting Grant	3,071	46.04%	2,332	51.19%	739	34.94%
Advanced Grant	2,142	32.11%	1,709	37.51%	433	20.47%
Consolidator Grant	959	14.38%	313	6.87%	646	30.54%
Proof of Concept	475	7.12%	178	3.91%	297	14.04%
Synergy Grant	24	0.36%	24	0.53%	0	0.00%
Total	6,671	100.00%	4,556	100.00%	2,115	100.00%

5.2 Distribution of scientific publications

We now report the distribution of the publications linked to ERC projects—namely, by focusing on distributions by ERC domain and by ERC project type. Table 5 shows the distribution of publications linked to FP7 and H2020 ERC grants by ERC sector and the number of publications per project. The average number of publications per project for the entire database was 25.89. The average publication-project time lag, as defined as the time lapse between the publication year of a publication and the starting year of its project, was 3.70 years. In terms of ERC sector, PE projects had larger publication portfolios in our dataset, with an average number of 34.10; by comparison, LS and SSH projects had, on average, 20.22 and 15.80 publications per project, respectively.

[11] Proof of Concept Grants (POC) are designed to encourage ERC grant winners to explore the innovation potential of the ideas developed under their previous ERC Frontier grants. In the analyses of this report presenting a breakdown of projects by scientific domains, we assign PoC projects to the scientific domain of their original frontier research project.

Table 5 - Distribution of ERC grants and related Scopus publications in the sample, by ERC sector of the project (all projects)

Domain	Total projects (a)	Total publications (b)	Average number of publications per project (b/a)
PE	3,077	104,930	34.10
LS	2,372	47,968	20.22
SSH	1,203	19,011	15.80
Synergy Grant	19	1,632	85.89
*Total**	*6,671*	*172,683*	*25.89*

*Note: * A publication can be linked to more than one ERC-project (i.e. to a PE project and to a LS project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique publications (reported in the final row of the table).*
Synergy Grant projects are usually multidisciplinary, spanning diverse scientific domains, so we report them as a separate category. However, for 5 Synergy grants we were able to identify a unique scientific domain of reference. These 5 projects are thus reported in this table in the specific scientific domain of reference

In terms of ERC project type (Table 6), AdG projects had larger publication portfolios, averaging about 36.64 publications per project, than either StG (24.17) or CoG (24.11). POC has few publications per project (1.92), which can be explained by the facts that (i) the majority of the publications are linked to the previous corresponding initial Frontier Research Grant and (ii) the shorter duration of PoCs (18 months) in respect to the other project types (5 years). The particular case of SyGs—programmes with more PIs working on the same project—showed the highest average number of publications per project (97.67).

Table 6 - Distribution of ERC grants and related Scopus publications in the sample, by ERC project type (all grants)

Project type	Total projects (a)	Total publications (b)	Mean number of publications per project (b/a)
StG	3,071	74,223	24.17
AdG	2,142	78,486	36.64
CoG	959	23,123	24.11
PoC	475	911	1.92
SyG	24	2,344	97.67
*Total**	*6,671*	*172,683*	*25.89*

*Note: * A publication can be linked to more than one ERC-project. For such reason, in this Table the sum of values related to projects of the various types is higher than the total number of unique publication (reported in the final row of the table).*

5.3 Distribution of patents (patents citing publications and self-reported patents)

We identified 34,513 patent applications citing scientific publications from ERC-funded projects included in our sample. In this section, we provide descriptive statistics on such patents, using information recovered through PATSTAT (Patstat Global edition Autumn 2021). Such patents cite a total of 12,871 scientific publications coming from 2,977 ERC-funded projects. On average, the share of projects cited by at least one patent application was equal to 44.63%. This percentage, however, varies significantly between the different ERC panels, as we detail in the subsequent section.

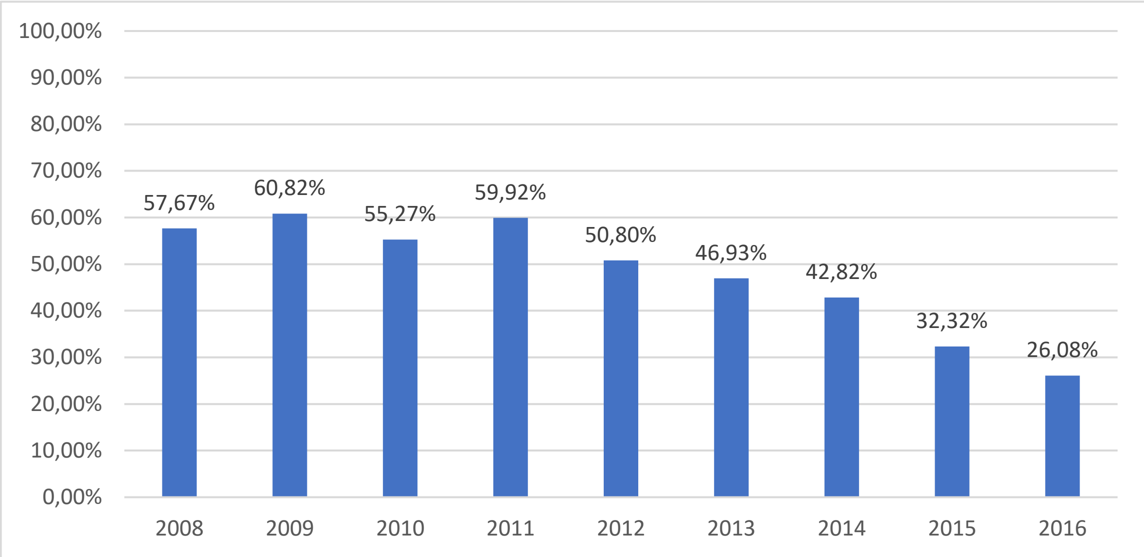
The 34,513 patents cited 12,871 publications (considering that one patent can cite more than one scientific publication stemming from ERC-funded research), out of a total of 172,683 publications linked to ERC-funded projects over the period of interest. Therefore, the percentage of scientific publications cited by subsequent patent applications was 7.45% in our dataset. This value is in the high-end of the range of average values reported by academic studies that have investigated papers cited in patents, which have been summarised in Section 3 of this report. In that Section, we have reported studies showing that, overall, 3-4% of papers are cited in patents when referring to papers from all scientific disciplines (as in the works of Van Raan, 2017 and de Moya-Anegon et al., 2020)[12]. However, it should be kept in mind that the results of such works are highly sensitive to a set of study-specific conditions (e.g., the type of scientific sectors considered; the data source used for identifying scientific publications and the related time coverage; the type of patents considered for the analyses; the time span considered to observe patents citing previous publication; the matching approach to link papers and patents). Thus, it is necessary to consider such average values as general reference points for interpreting the results, rather than precise threshold levels for assessing impact.

The large majority of citing patents are linked to FP7 projects (32,728 patents linked to 2,370 FP7 projects, corresponding to a 52% share of FP7 projects cited by at least one patent application). Not surprisingly thus, the number of H2020 projects (funded over the years 2014-2016) linked to subsequent patens is much lower (607 H2020 projects cited via publications by 3,204 patent applications, corresponding to a share of projects cited by at least one patent application equal to 28.7%), due to the restricted time window available for generating patent-to-publication citation process. In the case of FP7 ERC projects, the overall percentage of papers cited in patents is 8.63%, whereas it is 3.51% in the case of H2020 projects, as a consequence of time truncation. In this respect, Figure 5 shows that the likelihood of a project to be cited by a patent application is significantly higher for older projects compared to more recent ones. This is a consequence of the

[12] Studies focusing on scientific publications from science and engineering (Veugelers and Wang, 2019; Ahmadpoor and Jones, 2017) have generally found an overall average percentage of papers cited in patents of around 10-11%, with significant variations across scientific fields, types of papers and types of patents. Such values are in line with the one we discuss later in our analyses on projects from LS and PE.

considerable time lag between achieving research and innovation results, applying for patents, and undergoing the patent examination process.

Figure 5 - Share of ERC Grants cited by at least one patent application, by starting year of the project



If we consider granted citing patents, instead of applications, the numbers become lower. More precisely, we found 19,461 granted patents citing publications from 2,282 ERC projects included in our sample. In the case of granted patents, a project's probability of having its publication cited by a subsequent patent is equal to 34.21% (43% for FP7 projects, 15.32% for H2020 projects)[13]. 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. This statistic is naturally affected by the length of the examination process, which can take several years to reach a final grant decision. For this reason, a large share of the most recent patent applications included in our sample are still under examination.

Table 7 - Number of patents citing scientific publications from ERC-funded projects (FP7 vs H2020)

Variable	Total Dataset	FP7	H2020 (2014-2016)
ERC Projects	6,671	4,556	2,115
Patent applications citing linked publications*	34,513	32,728	3,204
Granted patents citing linked publications*	19,461	18,814	1,373

*Note: * A patent application can be reported by more than one ERC-project (i.e. by a FP7 project and by a H2020 project). For such reason, in this Table the sum of values related to FP7 and H2020 projects is higher than the total number of unique patent applications and granted patents (reported in the column "Total Dataset").*

[13] Despite the complexity of making direct comparisons due to different study conditions and types of data, it is interesting to notice that the overall percentage is similar to the one found in the study by Li et al. (2017), which considered research grants on biomedical research funded by the NIH in the US over a 27-year period. Those authors showed that about 30% of NIH grants generated articles that were subsequently cited by patents.

It can be useful to compare this first evidence to another benchmark in order to have a deeper insight on the ability of ERC-funded research to inspire subsequent technological developments. We now consider patented inventions that are direct outcomes of the ERC grants, based on self-reported project results disclosed at the Agency by the PIs. Table 8 reports similar information for reported patents. We were able to detect 1,550 unique self-reported patent applications, of which 507 resulted in granted patents at the time of this report (see Table 8). The vast majority of self-reported patents come from projects activated under the FP7 programme (1,340 patent applications, resulting in 459 granted patents at the time of the report), due to the lengthy process leading to publications and related citing patents.

In light of the fact that one project can generate multiple reported patents, 10.49% of the projects included in our dataset generated at least one self-reported patent application (700 projects out of 6,671 included in the sample).

Table 8 – Number of self-reported patents linked to FP7 and H2020 projects

Variable	Total Dataset	FP7 Projects	H2020 Projects (2014-2016)
Total number of ERC Projects	6,671	4,556	2,115
Number of ERC Projects with at least one self-reported patent applications	700	569	131
Self-reported Patent applications*	1,550	1,340	222
Self-reported granted patents*	507	459	51
Self-reported DOCDB Patent Families*	1,370	1,174	214
Self-reported INPADOC Patent Families*	1,353	1,159	212

*Note: * A patent application can be reported by more than one ERC-project (i.e. by a FP7 project and by a H2020 project). For such reason, in this Table the sum of values related to FP7 and H2020 projects is higher than the total number of unique patent applications, granted patents, DOCDB Patent Families and INPADOC Patent Families (reported in the column "Total Dataset").*

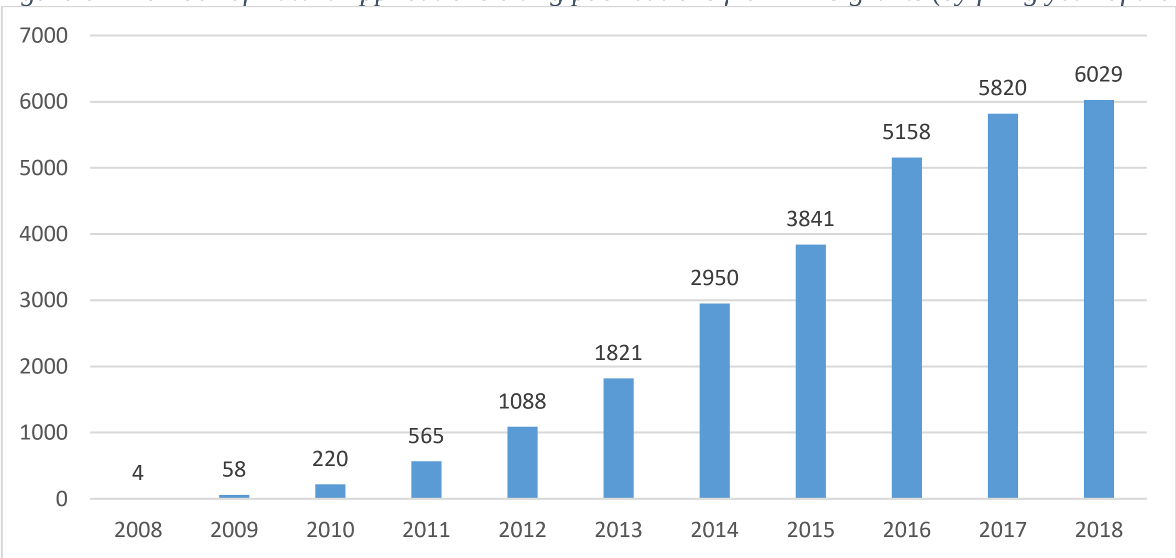
There is an immediately apparent difference in the number of patents we were able to link to ERC science funding via these two different approaches (patent-to-paper citations vs direct self-reporting). We showed a much larger set of ERC grants producing research that is cited by patents (around 44.6%) as compared to the set of grants directly reporting patents as project outcomes (around 10.4%). The method based on patents-publications-grants matches showed a higher percentage in terms of patent outcomes, as it captures patents that can be inspired by the projects' findings through indirect channels and even after several years from the end of the project itself, thus mapping influence even in the medium- and long-term.

A deeper analysis of the data revealed another interesting insight: namely, a limited overlap between the two groups of patents. Indeed, out of 1,550 self-reported patents, only 337 patents (around 22% in terms of share) also cited previous publications from the same project; thus, they were also included in the group of patents citing ERC-supported publications[14]. In terms of projects, 240 projects (for a share of 3.60%) declaring patents at the ERC as a direct outcome included a citation to previous publications stemming from the same project. We interpret this result as a byproduct of the research team deciding to first patent and then to publish in order to avoid undermining the novelty requirement for patenting (at least for the initial inventions generated as a result of the project).

Distribution of patents by year

Figure 6 below reports the distribution by filing year of patent applications citing ERC-related publications. In this respect, it should be noted that lag times between ERC initial funding and follow-on patenting via citation linkages are, on average, equal to 5.56 years in our dataset. That said, there are also patents citing publications from projects started 12 years before the filing year. The average lag between the filing year of the citing patent and the publication year of the cited publication in the NPL is 3.07 years in our dataset. This is due to the time lag required to achieve publishable research results and complete different patent applications based on such results. Moreover, any interpretation of these results needs to consider that patent applications are typically published by patent offices (and therefore made available to the public) after the expiration of a period of 18 months from the earliest filing date; until then, information on the patent remains secret. The reduction of patent numbers in the most recent years (2019, 2020, 2021) should therefore be interpreted as a result of incomplete disclosure on recent patent applications.

Figure 6 – Number of Patent Applications citing publications from ERC grants (by filing year of the patent)[15]

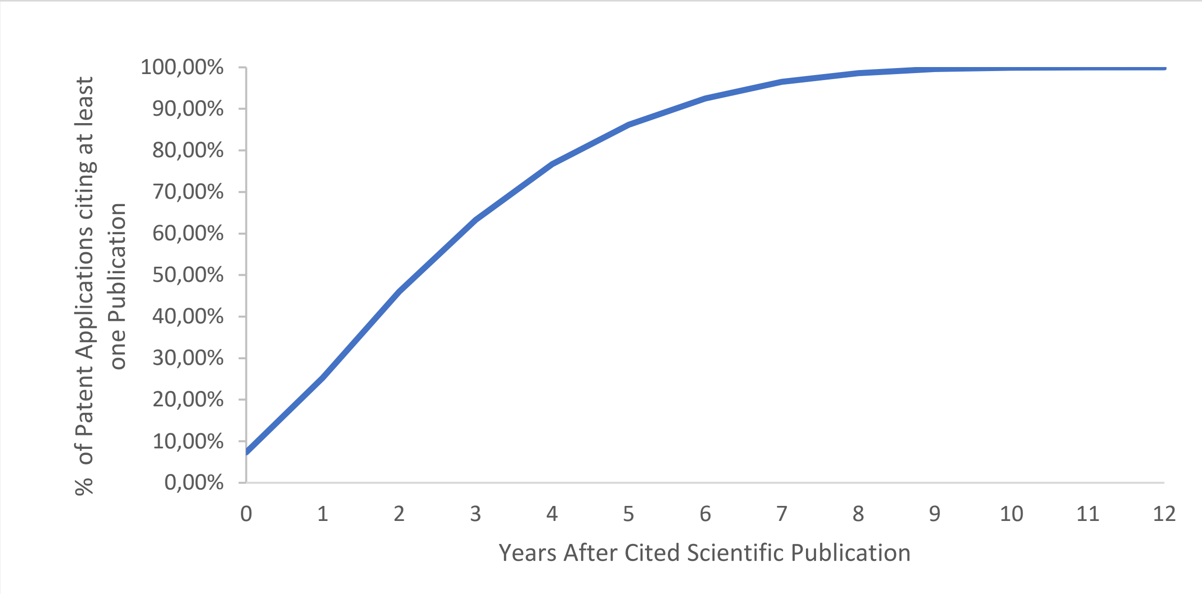


[14] In the group of self-reported patents, there are also 81 other patent applications citing publications linked to other ERC projects (different from the one generating the declared patent).

[15] We do not report years from 2019 to 2021 in this Figure due to the fact that patent applications are typically published by patent offices (and therefore they became publicly available) after the expiry of an 18-month secrecy period from

Figure 7 shows the distribution of patent applications according to the difference between their filing year and the year of the cited scientific publication. This provides a glimpse into how rapidly scientific knowledge can be translated into the patenting process. The figure shows that, although the average time lag is 3.69 years, there are also a few patents in the dataset citing publications published up to 10 years earlier.

Figure 7 – Cumulated distribution of patent applications, by time lag between the year of the cited scientific publication and the filing year of the patent



Distribution of patents by ERC sector

Looking at the distribution of citing patents in terms of ERC sector of the related projects in Table 9, we note that projects from the LS sector exhibit a higher probability of receiving at least one citation (through related publications) from subsequent patents (for LS projects, this probability is equal to 61.30%). The likelihood of being linked to a citing patent is lower in the case of PE projects (46.18%). As expected, the share of projects from the SSH sector being cited in patents is low, although not null, due the technical nature of the knowledge that is typically at the basis of patented inventions (the share of SSH projects linked to at least one citing patent is equal to 7.48%). The relative distribution across projects is very similar for FP7 projects as compared to H2020 projects, although the latter class exhibits lower percentage values in all classes.

the date of filing or the earliest priority date. For such reasons, the coverage of patents filed in the most recent years is limited. Considering that we used the version of Patstat released by the EPO in November 2021, the lower numbers of patent applications observed in the most recent years (2019, 2020, 2021) should be interpreted as a result of the incomplete disclosure of recent patent applications in this time window.

Table 9 – Number (and share) of projects cited by a patent application (via their publications), by ERC sector of the project (all projects in the sample)

Domain	Total projects (a)	Number of projects cited by a patent application (b)	Share of projects cited by at least one patent application (b/a)
LS	2,372	1,454	61.30%
PE	3,077	1,421	46.18%
SSH	1,203	90	7.48%
Synergy Grant	19	12	63.16%
Total	6,671	*2,977*	44.63%

Table 10 reports the breakdown of cited publications by ERC sector, showing significant variation across scientific fields in the patent-to-paper citation probability; this aligns with the results of previous literature in this area. The largest percentage of papers cited in patents, relative to the total number of papers assigned in that sector, comes from the LS sector (around 12%), followed by the PE sector (around 6.58%). Such percentages are typically higher if we focus solely on FP7 projects (respectively, 13.86% for LS projects and 7.61% for PE projects), whose publications have a longer time span for being cited compared to more recent H2020 projects[16].

Table 10 - Number and share of projects' publications cited by a patent application, by ERC sector of the project (all projects in the sample)

Sector	Total projects (a)	Total publications (b)	Number of publications cited by a patent application (d)	% Publications cited by a patent application (d/b)
LS	2,372	47,968	5,777	12.04%
PE	3,077	104,930	6,909	6.58%
SSH	1,203	19,011	177	0.93%
Synergy Grant	19	1,632	89	5.45%
Total*	6,671	172,683	12,871	7.45%

*Note: * A publication can be linked to more than one ERC-project (i.e to a FP7 and to a H2020 project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique publications (reported in the final row of the table).*

If we look at the similar distribution in the case of patents self-reported at the ERC by PIs, we observe slightly different results. LS and PE projects have a similar likelihood of generating at least one declared patented invention as a direct project outcome (13.45% of LS projects generated at least one self-reported patent, compared to 12.09% for PE projects). Such a percentage is definitively lower for SSH projects (0.4%), with only a few of them declaring a direct patent outcome. This result, when compared to the one found in the

[16] As previously reported, such values are in line with the ones obtained by previous studies looking at patent-to-paper citations in the fields of science and engineering (Ahmadpoor and Jones, 2017; Jefferson et al., 2018; Veugelers and Wang, 2019), which generally show an overall percentage of papers cited in patents of around 10-11%.

case of patent-to-paper citations (showing a significantly higher share in the case of LS), could probably be partly explained by the fact that patented inventions in the life sciences tend to be closer to the science-technology frontier (Ahmandpoor and Jones, 2017). This generally stems from the LS field needing more time for the innovation development process compared to other domains (Kordal et al., 2016).

Table 11 - Number and distribution of self-reported patent applications, by ERC sector of the project (all projects in the sample)

Sector	**Total projects** (a)	**Number of projects with at least one self-reported patent application** (b)	**Share of projects with at least one self-reported patent application** (b/a)
LS	2,372	319	13.45%
PE	3,077	372	12.09%
SSH	1,203	5	0.42%
Synergy	19	4	*21.05%*
Total	6,671	700	10.49%

Distribution of patents by family size

The protection conferred by patents is territorial: in order to protect a single invention in different geographic markets, inventors need to acquire a number of national or regional patents. A single invention, therefore, may be protected by a 'family' of patent documents across multiple markets. A patent family is defined as "the set of patents (or applications) filed in several countries which are related to each other by one or several common priority filings" (OECD, 2009). Patent families generally refer to the whole set of patents covering the same invention in one or more countries. The number of patent families thus indicates the number of distinct inventions being protected.

The average DOCDB family of the citing patents spans 3.67 jurisdictions, whereas the average DOCDB family for self-reported patents is higher, spanning 4.17 jurisdictions[17]. Figure 8 and Figure 9 respectively report the distribution of citing patents and self-reported patents by the number of legislations, for the DOCDB definition of patent families; the former distribution has a notably longer tail.

[17] A DOCDB family comprises all patent documents that have exactly the same priority date or a combination of priority dates and are related to the same invention

Figure 8 - Distribution of patent applications citing publications from ERC grants, by number of disclosed legislations in the respective DOCDB patent family

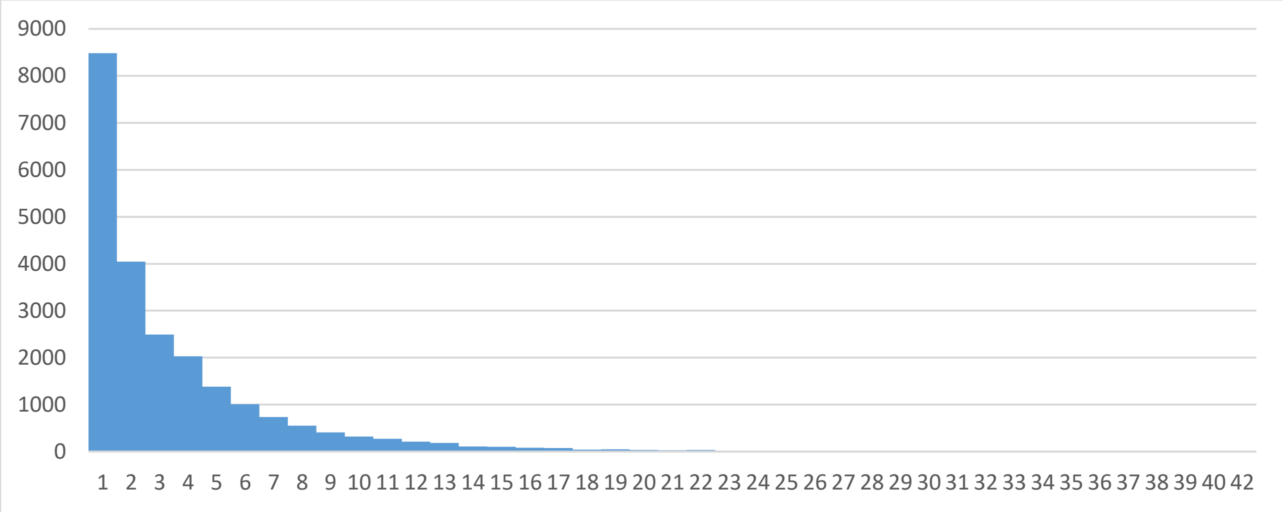


Figure 9 - Distribution of self-reported patent applications by number of disclosed legislations in the respective DOCDB patent family

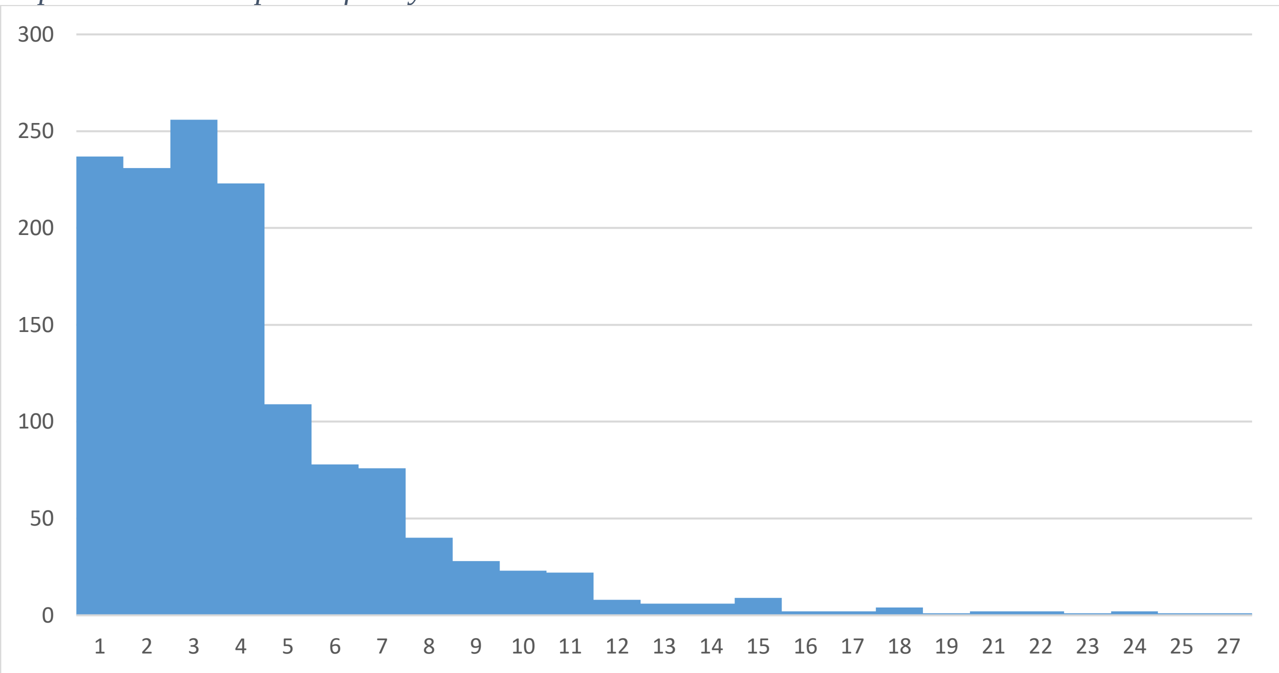


Figure 10 and Figure 11 report information about the size of the DOCDB patent families included in our dataset op patents citing ERC publications. The average patent family size ofpatents linked to ERC projects via cited publications is 6.7, but 5.14 in the case of self-reported patents.

Figure 10 - Distribution of patent applications citing publications from ERC grants by size of the respective DOCDB patent family

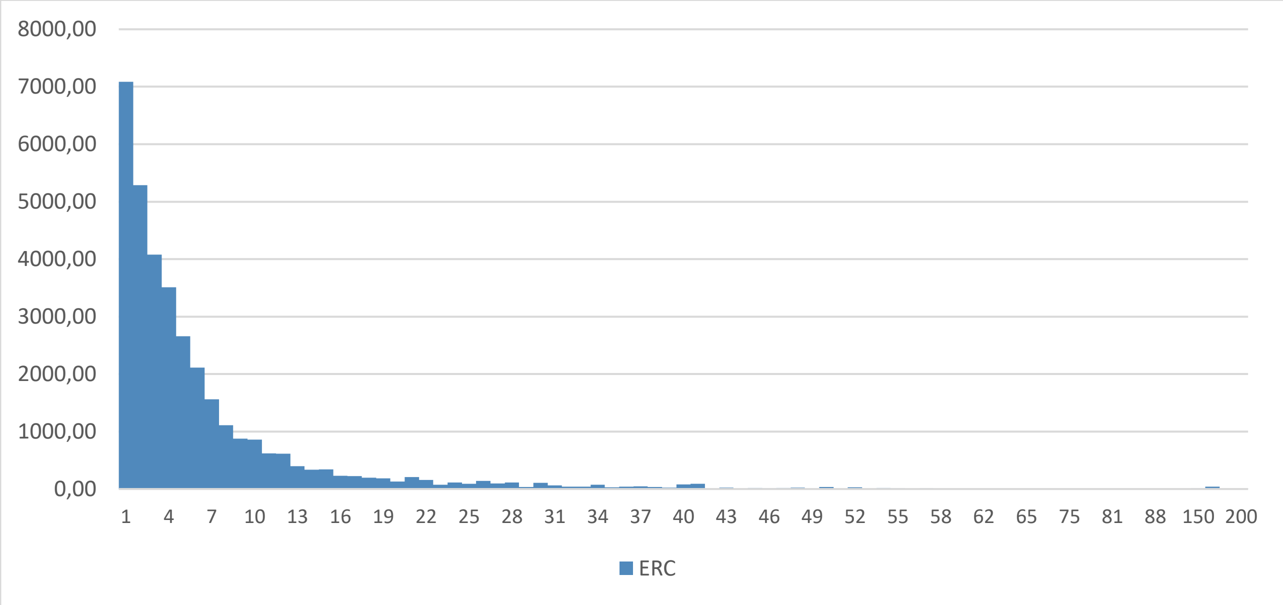
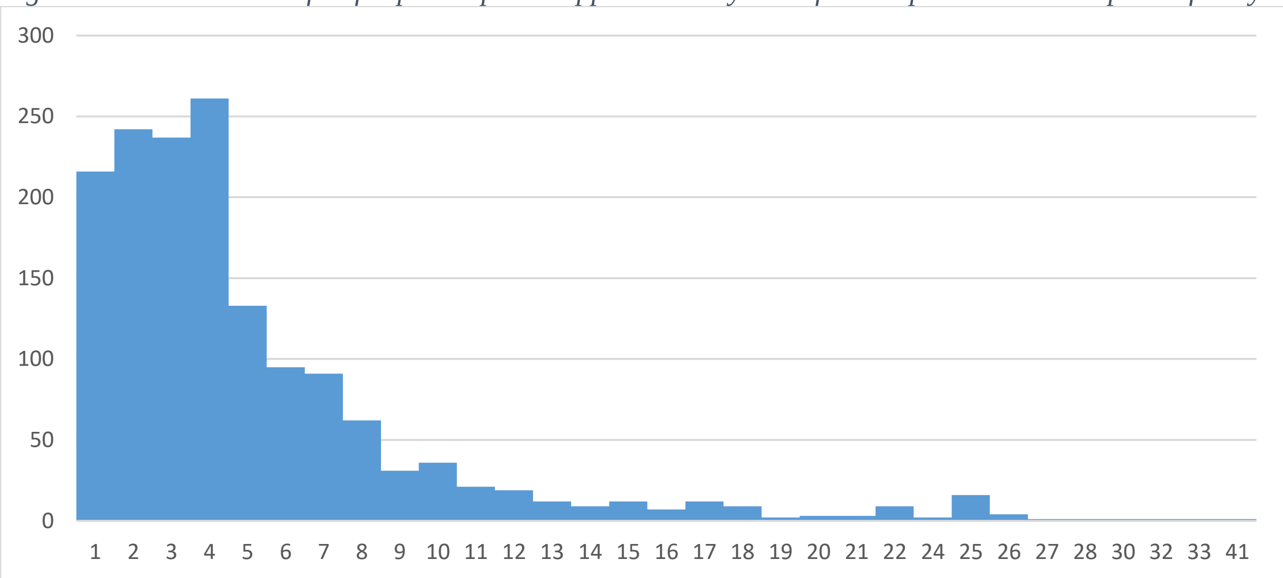


Figure 11 - Distribution of self-reported patent applications by size of the respective DOCDB patent family



Distribution of patents by type of applicant

We now turn to examine the breakdown of patents inspired by the type of applicant. The first analysis concerns the set of patents citing ERC-related publications: We want to understand how many of such patented inventions were generated by the same research team responsible for the project. As we better explain in the methodological section of the Annex, we sought to disentangle so-called "direct" patents from "indirect" patents in this sample of citing patents. According to our definition, the former are patents citing previous ERC-related publications that have the same Principal Investigator of the project (among the applicants and/or inventors) and the same Host Institution (Coordinator) or partner institution among the applicants. The assumption here is that such patents are directly connected to the scientific outcomes of the project. Adopting scientific publications as an analogy, these patents can be conceived as "self-citing". So-called "indirect" patents, meanwhile, are those assigned to different applicants and inventors; we consider them to be a spillover of the scientific knowledge generated through the project[18].

Our data show that only 2,057 patents (of 34,513 patent applications, or 5.96%) can be considered as directly coming from the project's research team. This percentage is slightly higher in the group of LS projects (7.21%) as compared to PE projects (4.71%). The large majority of such patent citations thus come from other institutional sources, showing a broader outreach of knowledge diffusion.

In order to better understand the institutions involved in patent inventions linked to ERC projects, we considered the five types of institutions that Patstat uses when categorising patent applicants: a) universities; b) research institutions (such as public research organisations); c) firms; d) individuals; e) other applicants (when not specified in Patsat). Table 12 below provides the breakdown of citing patent applications by type of institution (we adopted fractional counting in order to avoid double counting, in cases where patents have multiple applicants of a different nature). While the majority of citations come from patents assigned to firms (more than 50%), a significant share of citing patents derive from universities and research organisations. This finding aligns with the idea that patents assigned to universities and research laboratories are closer to the scientific frontier. The analysis also suggests that new knowledge stemming from ERC-funded research is a source of influence for additional technical development work that is still conducted in the scientific environment. Annex 2 at the end of this report reports the names of the first 20 firms with the highest number of patents citing, in their NPL section, previous ERC-funded research.

Table 12 - Distribution of patents citing ERC-funded publications by type of applicant and project scientific domain (fractional count), (all patents)

Domain	Total Citing Patent Applications	Firms	Universities	Research Institutes	Individuals	Other
LS	16,567	7,253.44	5,859.01	2,055.32	975.21	424.02
PE	18,303	10,249.70	5,467.07	1,306.68	955.36	324.19
SSH	389	263.98	60.37	26.20	29.30	9.15
Synergy Grant	284	135.35	84.64	43.67	10.71	9.63
Total*	***34,513***	***17,425.69***	***11,096.03***	***3,322.35***	***1,923.04***	***745.89***

*Note: * A patent application can cite publications linked to more than one ERC-funded project (e.g. to a LS and to a PE project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique patent application (reported in the final row of the table).*

[18] This selection procedure to identify Direct Patents is quite strict, but serves to avoid the problem of homonimy among PIs. An alternative procedure could be to use the PI as inventor/applicant or (instead of "and") the Host Institution as applicants of the papers. In this case, the number of direct patents is going to increase at the expense of possible errors in the perfect identification of the applicants related to the ERC projects.

If we consider the same type of breakdown for self-reported patents (Table 13), we note a different distribution: Universities (48% of cases) and research institutes (23% of cases) play a dominant role as patent applicants, coherently with a view of university or PRO institutional ownership in case of patents covering inventions to which academic research has contributed. In this group, patents owned by firms cover a smaller share (15% of all patents reported at the ERC), which signals and research and industry are engaging in collaborative efforts. Individuals are the applicants in a small share of self-reported patents (13%) (probably the academic inventors themselves).

Table 13 - Distribution of self-reported patents by type of applicant and project scientific domain (fractional count), (all patents)

Domain	Total Self-reported Patent Applications	Companies	Universities	Research Institutes	Individuals	Unknown
Life Sciences	648	84.41	308.22	171.33	77.40	6.64
Physical Engineering	885	149.78	422.59	176.07	119.53	17.03
Social Humanities	6	0.00	3.33	2.17	0.00	0.50
Synergy Grant	9	3.00	3.17	2.83	0.00	0.00
Total*	***1,548.00***	***237.19***	***737.30***	***352.40***	***196.93***	***24.18***

*Note: * A patent application can be reported by more than one ERC-funded project (e.g. by a LS and by a PE project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique patent application (reported in the final row of the table).*

6. The distribution of patents by technology domains

6.1 Analysing patents by technology domains: approaches

In this section, we analyse the technological diversity of patented inventions that have been linked to ERC-funded research. We include (1) patent applications that have cited scientific publications stemming from ERC projects as a source of inspiration, and (2) patents that have been directly generated from such projects, as reported by the PIs at the Agency. Various technological classifications of patents have been deployed and used over time, by both institutions and researchers. In this part of the report, we rely on the following classification approaches in order to better understand the influence exerted by ERC-funded research in various technological domains:

1) We first consider the distribution of citing and declared patents according to the WIPO classification of technological fields (35 fields) and related technological macro-sectors (5 sectors), using the WIPO IPC concordance table (WIPO, 2009)[19]. This classification has the advantage of being established in existing patent databases (e.g., Patstat, OECD statistics on patents), as well as highly adopted both in policy analyses (see, for instance, the OECD statistics by technology at the country level) and in the existing literature on patent-to-paper citations (i.e., Jefferson et al., 2018). The 5 macro-sectors and 35 related technological fields, as defined by the WIPO classification scheme, are reported below in Table 14. A more detailed description of the origins and underlying logic of this classification scheme (and of the foundational IPC classification) is reported in Annex 3 at the end of this report.

[19] The acronym IPC stands for "International Patent Classification", a hierarchical classification system used primarily to classify and search patent documents (patent applications, specifications of granted patents, utility models, etc.) according to the technical fields they represent. The Cooperative Patent Classification (CPC) is an extension of the IPC. See Annex 3 for more details.

Table 14 – The WIPO classification of patents by technology sectors and technology fields

Field_number	Technology Fields	Technology Sector
1	Electrical machinery, apparatus, energy	Electrical engineering
2	Audio-visual technology	Electrical engineering
3	Telecommunications	Electrical engineering
4	Digital communication	Electrical engineering
5	Basic communication processes	Electrical engineering
6	Computer technology	Electrical engineering
7	IT methods for management	Electrical engineering
8	Semiconductors	Electrical engineering
9	Optics	Instruments
10	Measurement	Instruments
11	Analysis of biological materials	Instruments
12	Control	Instruments
13	Medical technology	Instruments
14	Organic fine chemistry	Chemistry
15	Biotechnology	Chemistry
16	Pharmaceuticals	Chemistry
17	Macromolecular chemistry, polymers	Chemistry
18	Food chemistry	Chemistry
19	Basic materials chemistry	Chemistry
20	Materials, metallurgy	Chemistry
21	Surface technology, coating	Chemistry
22	Micro-structural and nanotechnology	Chemistry
23	Chemical engineering	Chemistry
24	Environmental technology	Chemistry
25	Handling	Mechanical engineering
26	Machine tools	Mechanical engineering
27	Engines, pumps, turbines	Mechanical engineering
28	Textile and paper machines	Mechanical engineering
29	Other special machines	Mechanical engineering
30	Thermal processes and apparatus	Mechanical engineering
31	Mechanical elements	Mechanical engineering
32	Transport	Mechanical engineering
33	Furniture, games	Other fields
34	Other consumer goods	Other fields
35	Civil engineering	Other fields

2) We then focus on two sets of technological domains that are particularly relevant for industrial development and policy-making, due to their widespread diffusion across a variety of industrial sectors and their tackling of pressing societal challenges. We refer to patented inventions related to climate change mitigation technologies, as well as patents related to the digital transformation. Both the green transition and the digital transition represent strategic priorities for the future of Europe in general, and the EU political and industrial agenda in particular. In the former case, we refer to the concordance schemes developed by the EPO (Angelucci et al., 2018), which introduced a dedicated tagging scheme known as the "Y02/Y04S scheme" (and related sub-fields) that is fully integrated within the CPC. This scheme covers seven main categories of climate change mitigation technologies, namely related to energy, greenhouse gases, buildings, industry and agriculture, transport, waste management, wastewater treatment, and smart grids. Table 15 below reports the descriptions of such fields. Annex 3 at the end of this report provides further details on this classification scheme.

Table 15 – The EPO classification of patents related to climate change mitigation and adaptation technologies (Angelucci et al., 2018)

CPC Classification	
Y02A	Technologies for Adaptation to Climate Change
Y02E	Reduction of greenhouse gas (GHG) emissions, related to energy generation, transmission or distribution
Y02P	Climate change mitigation technologies in the production or processing of goods
Y02T	Climate change mitigation technologies related to transportation
Y02B	Climate change mitigation technologies related to buildings
Y02C	Capture, storage, sequestration or disposal of greenhouse gases (GHG)
Y02D	Climate change mitigation technologies in information and communication technologies (ICT)
Y02W	Climate change mitigation technologies related to wastewater treatment or waste management
Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e., smart grids

For what concerns technologies related to the digital transition, we refer to the concordance scheme developed by the EPO in the study on patents related to the Fourth Industrial Revolution (EPO, 2017). Also in this case, a concordance table linking together specific 4IR technology sectors and a set of corresponding CPC patent classes is available. 4IR patented inventions have been classified by the EPO into three main sectors (Core Technologies, Enabling Technologies, Application Domains), each of which is subdivided into several technology fields (for a total of 18 fields), as reported in Table16.

Table 16 - The EPO classification of patents related to 4th Industrial Revolution technologies (EPO, 2017)

Technology Sector	Field	Definition
Core Technology	Connectivity	Network protocols for massively connected devices, adaptive wireless data systems for short-range and long-range communication
	IT Hardware	Sensors, advanced memories, processors, adaptive displays, smart instruments
	Software	Intelligent cloud storage and computing structures, adaptive databases, mobile operating systems, virtualisation and blockchain technologies
Enabling Technology	Core Artificial Intelligence	Machine learning, neural networks, statistical and rule-based systems, AI platforms
	Data Management	Diagnostic and analytical systems for massive data, prediction and forecasting techniques, monitoring functions, planning and control systems
	Data Security	Adaptive security systems for devices, services and data transmission
	Geo-Positioning	Enhanced geo-location and satellite navigation, device to device relative and absolute positioning
	Power Supply	Automated generation, situation-aware charging systems, shared power transmission and storage objectives, smart power-saving management
	Safety	Intelligent safety systems for theft and failure prevention
	Three-dimensional support systems	3D printers and scanners for parts manufacture, automated 3D design and simulation, 3D user interfaces
	User Interfaces	Virtual reality, augmented reality, speech recognition and synthesis
Application Domains	Agriculture	Climate monitoring systems, greenhouse automation, smart crop and cattle management, smart farming
	Consumer Goods	Personal health monitoring devices, smart wearables, smart entertainment and sport devices, smart toys and textiles
	Healthcare	Intelligent healthcare systems, robotic surgery, smart diagnosis
	Home	Smart homes, alarm systems, intelligent lighting and heating, consumer robotics, climate control systems
	Industrial	Smart factories, intelligent robotics, energy saving
	Infra-Structure	Intelligent energy distribution networks, intelligent transport networks, intelligent lighting and heating systems
	Services	Intelligent retail, payment and loyalty systems, smart offices
	Vehicles	Autonomous driving, vehicle fleet navigation devices

3) Finally, we performed a detailed analysis on the set of ERC FP7 and H2020 grants that received the highest number of citations by subsequent patents. By identifying the respective technological domains, such analyses illustrate the projects that are particularly influential in inspiring subsequent technological developments.

In the following section, we report a series of analyses on the patents citing ERC-funded projects based on their technological domain. For a general overview, we follow the standard OECD approach in technologically classifying patent statistics. More specifically, we rely on the WIPO concordance table based on the four-digit IPC patent classes to derive a classification of technological fields (35 fields) and their higher aggregation in technological macro-sectors (5 sectors). We then use the concordance schemes developed by the European Patent Office to focus on a) climate change mitigation technologies and b) fourth industrial revolution technologies. For any given classification, every cited ERC project is then associated with the class where the corresponding citing patent falls; fractional counts are used to accommodate multiple classifications[20].

[20] If one patent application is assigned to more than one technology sector (due to the presence of multiple IPC classes in the application), it can either be partly attributed pro quota to each technology sector involved (fractional counts) or fully attributed to each sector (full counts). With the fractional count approach, the application is divided equally among all the technology sectors involved, thus avoiding double counting in the analyses. We adopt this approach in our analyses, as it is the same used by the OECD in compiling its statistics of patents by technology at the country level.

6.2 Distribution of patents by WIPO technological fields

We start our analyses by focusing on the technological fields of reference among patents citing publications (in their NPL section) from ERC-funded projects. We used the technology classification groups of the WIPO concordance tables, which links IPC classes with 5 major technological sectors and their related 35 fields of technology. Starting from the more general aggregation, we notice that patents relying on ERC-funded publications are mostly coming from the “Chemistry” (52.11%) and “Electrical Engineering” (28.69%) WIPO technological macro-sectors. The “Instruments” sector follows (17.19%), while the “Mechanical Engineering” (1.63%) and “Other sectors” (0.33%) show negligible percentages. Figure 12 reports these values. If we look at their split between FP7 and H2020 projects, while the overall pattern holds for both framework programs, it is not able that the Mechanical Engineering class increases its value in the H2020 projects (2.48% vs. 1.62%), although it remains significantly lower than the other three preceding it.

Figure 12 – Distribution of patent applications citing publications from ERC grants, by WIPO macro technological sectors (in % of total patents; fractional count; all patents)

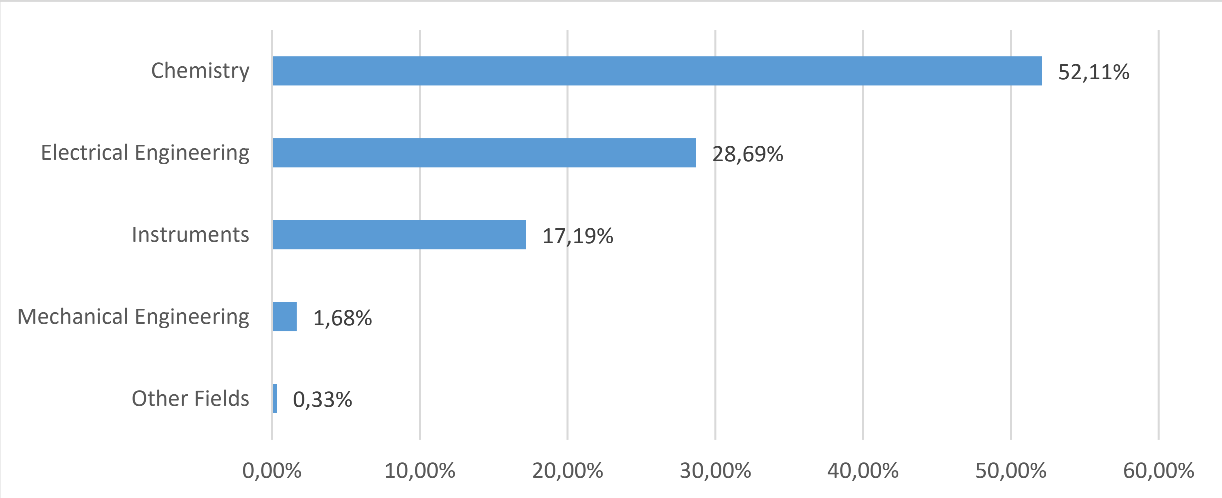


Table 17 reports the data distinguishing between projects from the LS and the PE ERC sectors, which together account for 98.9% of all citing patent applications. As we expected for both the whole dataset and for the FP7 and H2020 programmes separately, the LS ERC-funded projects are mainly cited in the Chemistry domain, which account for 81.04% of all citations. On the other hand, PE ERC-funded projects show a more diversified impact: While they are mainly cited in the Electrical Engineering domain, they are also highly cited by the Chemistry (26.77%) and Instruments (20.09%) domains. A closer look also reveals that the growth of the Mechanical Engineering domain is mainly due to H2020 ERC-funded projects in the PE field.

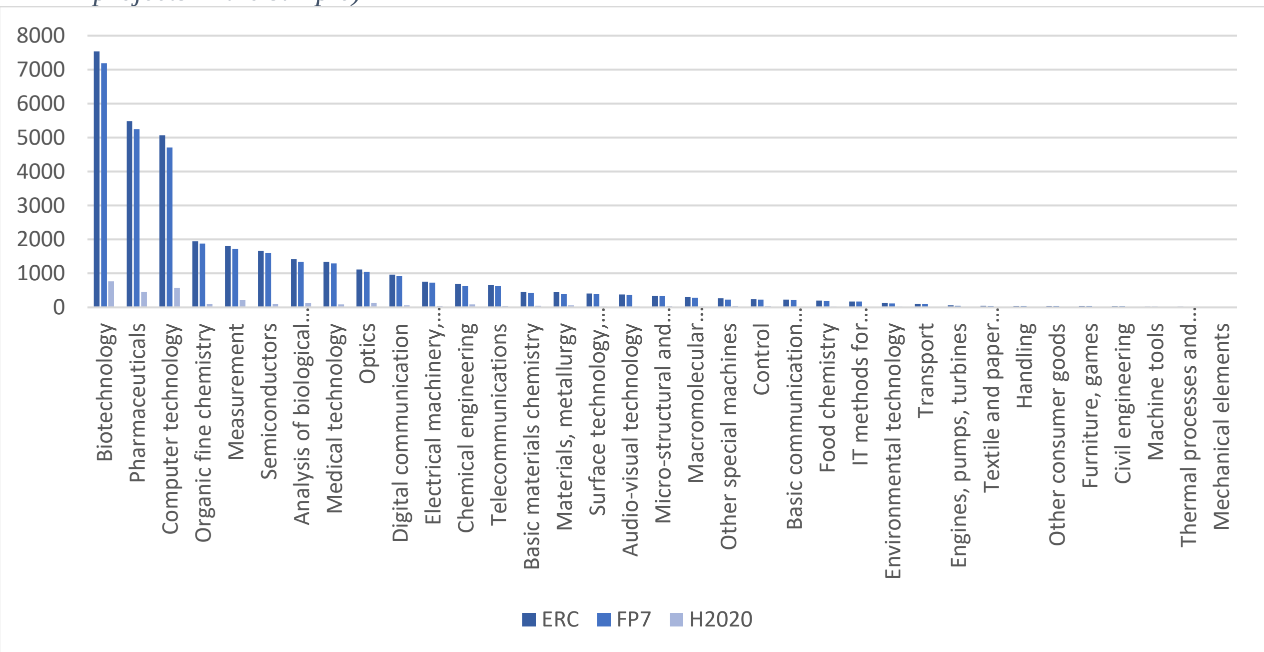
Table 17 - Distribution of citing patents by ERC sectors and WIPO technological-macro sectors (in % of all patents linked to projects of such scientific areas; fractional count)

WIPO Technology Field	Life Science (LS) Projects		Physical Engineering (PE) projects	
	Number of patent applications	Share (%) of patent applications	Number of patent applications	Share (%) of patent applications
Electrical Engineering	648.87	3.92%	9,166.17	50.15%
Instruments	2,334.19	14.11%	3,671.36	20.09%
Chemistry	13,409.69	81.04%	4,893.06	26.77%
Mechanical Engineering	136.78	0.83%	451.45	2.47%
Other fields	17.47	0.11%	96.96	0.53%
Total*	***16,547.00***	***100.00%***	***18,279.00***	***100.00%***

*Note: * A patent application can be reported by more than one ERC-funded project (e.g. to a LS and to a PE project). For such reason, in this Table the sum of values related to patent applications from the various domains is higher than the total number of unique patent applications.*

Figure 13 reports the distribution in all 35 technological classes, distinguishing between the overall numbers of projects cited and the programme-specific counts. The distribution is evidently rather skewed, with over 50% of citations concentrated in three classes and the remaining 50% distributed among the other thirty-two. Those three classes—namely, Biotechnology, Pharmaceuticals and Computer Technologies—and their concentration do not change between programmes, although Computer Technologies increases its weight from 14.4% in FP7 to 18.8% in H2020, which contributes to increasing the programme's overall weight. The Figure shows that patent citations to ERC-based research outputs are more common in technologies closest to the science frontier and in areas where industry has a heavy science-based R&D orientation. This evidence is in line with previous studies on patent-to-paper citations using different contexts of analyses (Ahmadpoor and Jones, 2017; de Moya-Anegon et al., 2020; Jefferson et al., 2018).

Figure 13 - Distribution of citing patents in the 35 WIPO technological fields (fractional count; all projects in the sample)



In order to have a further, more granular, overview of technologies impacted by ERC-funded research, Table 18 reports the twenty most frequent IPC subclasses (4-digit level) observed in the dataset of citing patent applications. On average, citing patents are assigned to 2.2 distinct IPC subclasses. Patent applications citing ERC publications cover a wide spectrum of technological areas (they are classified in 412 IPC subclasses out of 647 subclasses existing in the current IPC scheme).[21] However, the distribution is skewed, with a very long tail, given that, as shown in Table 18, the first eight IPC subclasses include 50% of the patent applications reported in the dataset.

Table 18 - Top 20 IPC Subclass more frequently on patent applications citing a ERC project

IPC Subclass	Description	Number of Patent Applications (Weight)	%	Acum
A61K	PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES	3801.12	11.03%	11.03%
C12N	MICROORGANISMS OR ENZYMES; COMPOSITIONS THEREOF; PROPAGATING, PRESERVING, OR MAINTAINING MICROORGANISMS; MUTATION OR GENETIC ENGINEERING; CULTURE MEDIA	3384.24	9.82%	20.85%
G01N	INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES	2348.50	6.81%	27.66%
C12Q	MEASURING OR TESTING PROCESSES INVOLVING ENZYMES, NUCLEIC ACIDS OR MICROORGANISMS (immunoassay G01N 33/53); COMPOSITIONS OR TEST PAPERS THEREFOR; PROCESSES OF PREPARING SUCH COMPOSITIONS; CONDITION-RESPONSIVE CONTROL IN MICROBIOLOGICAL OR ENZYMOLOGICAL PROCESSES	1927.85	5.59%	33.25%
C07K	PEPTIDES	1680.12	4.87%	38.13%
G06F	ELECTRIC DIGITAL DATA PROCESSING	1520.08	4.41%	42.54%
A61P	SPECIFIC THERAPEUTIC ACTIVITY OF CHEMICAL COMPOUNDS OR MEDICINAL PREPARATIONS	1513.52	4.39%	46.93%
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR	1494.55	4.34%	51.26%
G06T	IMAGE DATA PROCESSING OR GENERATION, IN GENERAL	1245.86	3.61%	54.88%
C07D	HETEROCYCLIC COMPOUNDS	941.57	2.73%	57.61%
G06N	COMPUTING ARRANGEMENTS BASED ON SPECIFIC COMPUTATIONAL MODELS	920.73	2.67%	60.28%
G06K	GRAPHICAL DATA READING (image or video recognition or understanding G06V); PRESENTATION OF DATA; RECORD CARRIERS; HANDLING RECORD CARRIERS	889.11	2.58%	62.86%
H04L	TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION	739.10	2.14%	65.00%
A61B	DIAGNOSIS; SURGERY; IDENTIFICATION	682.96	1.98%	66.99%
G02B	OPTICAL ELEMENTS, SYSTEMS OR APPARATUS	462.44	1.34%	68.33%
H04B	TRANSMISSION	369.45	1.07%	69.40%
B82Y	SPECIFIC USES OR APPLICATIONS OF NANOSTRUCTURES; MEASUREMENT OR ANALYSIS OF NANOSTRUCTURES; MANUFACTURE OR TREATMENT OF NANOSTRUCTURES	353.57	1.03%	70.42%
C12P	FERMENTATION OR ENZYME-USING PROCESSES TO SYNTHESISE A DESIRED CHEMICAL COMPOUND OR COMPOSITION OR TO SEPARATE OPTICAL ISOMERS FROM A RACEMIC MIXTURE	344.70	1.00%	71.42%
B01J	CHEMICAL OR PHYSICAL PROCESSES, e.g. CATALYSIS OR COLLOID CHEMISTRY; THEIR RELEVANT APPARATUS	327.00	0.95%	72.37%
C07C	ACYCLIC OR CARBOCYCLIC COMPOUNDS	304.37	0.88%	73.26%
Total		34469.00	100.00%	

[21] According to the World Intelectual Property Organization (WIPO), there are 647 IPC Subclasses. Information available on: https://www.wipo.int/classifications/ipc/en/ITsupport/Version20220101/transformations/stats.html .

In Table 19 (whole sample), Table 20 (FP7 Grants), and Table 21 (H2020 Grants) we further analyse the differences among ERC domains. In the tables, the rows list the programmes' subsectors (panels) and the columns list the 35 WIPO Technology Fields. Each cell reports the total number of citing patent applications. The matrices use the intuitions developed by Cohen, Nelson, and Walsh (2002), and further discussed by Mazzoleni and Nelson (2007), to map knowledge flows from science developed within the ERC-funded projects and the generation of new technologies. While Cohen et al. (2002) based their observations on R&D managers' perceptions of the relative importance of different fields, we rely on a more direct and robust measure. Specifically, we use the patent citation received by the scientific papers published by the different ERC sectors which are classified according to their respectively WIPO categories. While we were not able to exclude the patent applications where the ERC grantees appeared as inventors or co-inventors (given the large number of citing patent applications), we can assume that their presence in the sample does not significantly impact the overall results.

In these three tables, each cell reports the number of patents in any given WIPO category that cited papers published out of projects financed in the different scientific fields. We can see from the matrices that some cells are empty or have a low frequency of patents, demonstrating a low interaction, while others report higher numbers. We highlight all cells that account for 10% of all citing patents in each of the four ERC macro scientific aggregations (LS, PE, SSH, multidisciplinary projects based on Synergy grants). Different cut off values do not show significantly different patterns.

Cohen et al. (2002) showed that applied science fields like Engineering, Computer Science, and Material Science presented a more widespread impact on industrial R&D, while Medical and Health Sciences contributed mainly to Drug Development and Medical Equipment. Basic sciences presented different results: Biology and physics were important to specific industries (pharmaceutical and semiconductors, respectively), mathematics did not have a clear impact on any field, and chemistry had an impact on a broader range of industries (i.e., food, petroleum, metals, drugs, and chemicals). Our matrices show a similar pattern of concentration based on disciplinary differences, although the fields are not perfectly comparable with those used by Cohen et al. (2002). The impact of LS is more concentrated and related to new technologies developed in the Chemistry WIPO category (biotechnology, pharmaceuticals, and organic fine chemistry) and the analysis of biological materials. Relatedly, ERC-LS projects grouping medical technologies, diagnostics, therapies, and public health show a high impact on patent applications in the medical technology sector (sector 13).

Table 19 - Distribution of citing patents by ERC panel (row) and WIPO Technology Fields (column), All grants – Nr of Citing Patent Applications in cell

Sector	Subsector	N Projects	Electrical Engineering								Instruments					Chemistry											Mechanical Engineering								Other Fields		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Life Sciences	LS1	169	15.02	1.36	0.67	0.11	0.62	41.32	0.08	2.45	14.44	45.98	166.07	0.28	14.91	**270.91**	**1064.14**	**642.91**	3.95	20.94	35.01		4.77	17.01	13.16		0.40		3.10	0.67	12.12	0.08			0.39	1.13	
	LS2	177	38.09	1.26	0.36	0.61	2.26	193.55		5.90	23.51	65.85	**206.01**	0.78	27.79	160.86	**1765.87**	**557.02**	1.35	42.35	31.81	2.96	1.00	2.58	34.39	2.11	0.76		0.11		19.87						
	LS3	128	0.25	0.96				20.28	0.33	0.20	26.73	18.14	103.01	0.40	19.83	81.61	**720.49**	**543.82**	1.68	15.62	8.39	0.14	0.46	4.41	20.68	3.05	0.98		1.00		15.93				1.40		0.20
	LS4	188	3.17	1.02	0.14	0.50	0.17	50.66	0.40	1.25	3.23	30.67	**221.51**	0.76	85.42	**276.05**	**841.97**	**1236.67**	1.51	22.97	11.17	2.39	1.05	2.67	8.76	0.19			0.34	0.60	14.41		0.22	0.42	1.62	0.09	
	LS5	160	2.65	3.78	1.90	10.17	0.25	64.80	4.10	1.87	39.23	27.03	55.06	3.64	161.39	71.56	**415.22**	**399.66**	1.50	2.04	10.03	0.83	2.23	2.70	2.76		0.93	1.07			15.14			0.17	1.28	0.25	3.76
	LS6	191	0.86		0.36	3.11		31.59	2.09		0.17	15.61	197.17	1.29	38.65	191.15	**1234.96**	**1253.60**	3.18	13.79	10.38	2.24	0.44	4.38	5.44	0.10	0.53	0.18			23.56		0.17				
	LS7	254	9.70	17.10	1.74	1.93	1.24	122.09	1.26	4.52	45.63	198.81	**271.24**	4.99	**335.28**	**249.57**	**1519.49**	**1240.36**	20.23	21.40	20.91	1.87	4.67	15.18	30.27	7.00	1.59	0.54	6.72	0.70	26.01		0.29		1.56	0.11	
	LS8	72	0.50			0.25		37.86	0.50			8.87	12.10	0.30	4.44	10.70	**201.12**	63.59	1.45	14.69	5.81			0.99	1.67	8.85					3.45			0.20		0.67	
	LS9	114	12.23	0.22	0.53		0.83	18.96		2.37	4.27	41.85	65.86	0.87	29.47	37.11	**939.76**	**251.37**	4.24	88.20	23.53	5.29	2.64	16.13	14.95	9.70	0.73	0.20	0.33	4.72	19.08					6.55	
Physical Engineering	PE1	24	2.52	8.37	1.33	2.85	1.53	82.38	2.50	0.80	14.04	17.68		1.80	16.03		8.17	2.98			2.00			0.13		2.26	1.45	0.73	0.46								
	PE2	143	33.67	16.02	51.70	30.28	8.18	153.85	3.78	33.43	**288.19**	**244.49**	3.04	5.65	26.68	1.58	21.89	2.70			4.10	2.54	5.27	18.35	12.35	10.61	2.00	1.25	9.71	0.89	3.50	1.00		2.76		0.68	1.85
	PE3	189	**184.24**	30.45	177.41	49.75	39.84	**263.61**		**814.23**	**253.41**	**397.19**	94.85	10.50	90.98	42.05	**184.25**	35.01	17.84	2.64	32.13	80.98	129.44	93.88	113.42	23.27	1.77	7.07	17.02	4.01	17.89	6.38	2.88	6.01	1.40	8.59	2.60
	PE4	159	126.93	5.02	4.51	0.14	2.82	43.83		123.55	81.25	**232.41**	127.76	1.20	59.36	**189.16**	**334.52**	158.40	40.90	2.53	77.55	106.88	45.18	49.96	180.30	27.80	0.94	1.70	4.25	3.00	8.31	1.58	6.08	1.03		0.29	2.85
	PE5	285	**266.63**	53.27	8.91	108.60	3.76	**364.33**	21.42	**430.34**	79.29	134.50	73.06	10.44	96.40	**601.96**	**238.03**	**277.07**	180.36	2.65	**190.18**	**185.79**	160.64	107.71	**184.60**	34.29	1.42	3.94	4.67	15.65	32.96	3.29	2.30	3.53	12.58	6.43	5.01
	PE6	221	44.70	**185.05**	31.32	**472.23**	19.71	**3262.70**	119.19	147.44	83.83	151.17	49.41	153.04	153.42	11.98	90.97	44.60	5.36	0.46	11.04	44.49	22.58	22.53	26.84	7.15	11.13	2.37	3.91	5.72	33.40	1.63	0.64	55.35	17.71	16.94	2.00
	PE7	172	74.78	77.28	**382.68**	**328.56**	149.03	**464.95**	9.33	**196.55**	**339.65**	**326.56**	15.73	33.61	160.25	2.35	48.67	7.16	6.44	0.33	6.33	3.75	18.28	28.92	13.26	0.83	16.47	2.81	8.87	0.42	19.19	0.20	1.73	18.59	1.78	0.68	0.96
	PE8	167	66.05	14.74	2.53	1.70	3.20	52.86	2.13	154.85	53.40	118.28	35.43	15.16	99.28	33.82	**209.10**	62.32	39.39	2.97	39.94	60.29	88.41	26.36	104.40	11.38	8.32	4.88	9.28	14.62	50.74	5.14	2.78	12.44	3.05	2.31	5.45
	PE9	24		0.45	4.33		0.33	10.77		1.97	1.75	11.40	0.33	1.00	1.50	0.50	4.00	0.50		0.67					3.00	1.00								0.50			
	PE10	36	8.06	0.50		0.85	0.33	13.51		0.60	2.98	34.82	6.04	1.27	2.02	4.10	12.28	3.99	2.97	0.76	6.49	0.08	1.00	0.75	2.30	3.25			4.65		2.53	2.45	0.33	7.66		1.00	6.42
Social Humanities	SH1	3						1.83	0.33					0.83			2.00																				
	SH2	4	0.40			1.67		0.33						0.20			1.00																	0.40			
	SH3	16	2.10	1.89	1.18	3.52	1.63	**19.96**	8.61		0.10	3.49		4.10	**42.54**	0.33	2.08	3.56		0.50				0.50		1.38			0.25		0.24		0.83	0.20			
	SH4	55	0.33	2.78	0.21	1.58		**94.55**	10.67		1.17	3.82	7.34	2.39	**66.99**	1.68	**23.26**	**11.62**		1.32	0.25						0.29							1.95	1.17	0.25	1.38
	SH5	4			0.07			3.86						0.07		0.50	1.00	0.50																			
	SH6	8				0.67		11.37	0.33	0.50	0.50	0.38	1.25	0.83		1.75	**23.78**				0.50	0.50			0.45						0.20						
Synergy Grant		12	1.86		2.90	0.75	1.87	29.57	0.50	2.82	5.57	**67.53**	25.05	2.04	12.31	6.45	**59.12**	52.16	1.00	2.00	2.76	0.58	0.06	1.72	3.83						0.33			0.88			0.33
Total Patent Applications by field*			761.20	379.06	651.52	965.53	225.42	5065.69	177.33	1661.99	1114.22	1808.43	1419.12	241.61	1341.09	1943.21	7535.37	5487.26	309.01	202.73	457.76	449.10	404.50	346.33	692.73	132.59	45.50	23.81	60.29	46.79	264.80	17.86	16.02	103.79	38.98	45.32	31.07

Note: * A patent application can cite publications linked to more than one ERC-funded project (e.g. to a LS1 and to a PE2 project). For such reason, in this Table the sum of values related to patent applications from the various technology field is higher than the total number of unique patent applications (reported in the final row of the table).

Table 20 - Distribution of citing patents by ERC sub-sector (row) and WIPO Technology Fields (column), FP7 grants – Nr of Citing Patent Applications in cell

Sector	Subsector	N Projects	Electrical Engineering								Instruments					Chemistry											Mechanical Engineering								Other Fields		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Life Sciences	LS1	143	15.02	1.36	0.67	0.11	0.62	40.70	0.08	1.78	13.78	44.74	147.35	0.28	14.55	**266.26**	**1024.42**	**614.23**	2.58	20.29	35.01		4.77	17.01	13.02		0.40		3.10	0.67	11.62	0.08			0.39	1.13	
	LS2	144	38.09	1.26	0.36	0.61	2.26	**187.84**		5.90	23.51	64.85	**198.45**	0.78	26.21	159.38	**1698.30**	**536.69**	1.35	41.85	30.55	2.96	1.00	2.58	34.39	2.11	0.76		0.11		19.87						
	LS3	110	0.25	0.58				18.28	0.33	0.20	21.86	14.58	96.38	0.40	18.97	80.78	**698.56**	**534.36**	1.68	15.62	8.39	0.14	0.46	3.27	19.68	2.72	0.98		1.00		14.93				1.40		0.20
	LS4	152	3.17	1.02	0.14	0.50	0.17	47.66	0.40	1.25	3.23	30.42	**212.74**	0.76	77.49	**267.56**	**802.65**	**1192.04**	1.51	22.97	11.17	2.39	0.08	2.67	8.76	0.19			0.14		11.80		0.14	0.42	1.47	0.09	
	LS5	136	2.65	3.56	1.90	9.17	0.25	57.58	4.10	1.62	38.57	25.25	53.12	3.64	158.39	68.01	**407.66**	**382.61**	1.50	2.04	10.03	0.83	1.98	2.70	2.26		0.93	1.07			15.14			0.17	1.28	0.25	3.76
	LS6	156	0.86					29.61	2.09		0.17	12.74	186.55	1.29	38.49	181.42	**979.57**	**1187.76**	3.18	9.74	9.46	2.24	0.44	4.24	3.44	0.10	0.53	0.18			19.75		0.17				
	LS7	178	9.70	14.43	1.23	1.93	1.24	101.53	1.26	4.52	45.00	183.13	**246.64**	4.50	**305.16**	**230.43**	**1420.15**	**1133.30**	20.23	19.34	20.16	1.87	4.67	15.18	26.83	7.00	1.59	0.54	5.72	0.70	22.60		0.15		0.15	0.11	
	LS8	61	0.50			0.25		35.86	0.50			8.87	12.10	0.30	4.44	9.50	185.81	55.72	1.45	10.13	4.71			0.99	1.67	8.85					1.50			0.20		0.67	
	LS9	86	10.36	0.22	0.53		0.83	17.46		2.37	4.27	40.43	62.95	0.87	28.64	34.76	**883.71**	**248.40**	2.62	82.81	21.20	5.29	2.64	15.29	12.78	8.37	0.73	0.20	0.33	4.72	18.66					6.55	
Physical Engineering	PE1	24	2.52	8.37	1.33	2.85	1.53	82.38	2.50	0.80	14.04	17.68		1.80	16.03		8.17	2.98			2.00			0.13		2.26	1.45	0.73	0.46								
	PE2	112	29.76	13.29	46.37	30.28	7.93	141.80	3.78	32.33	**266.08**	**228.66**	2.54	5.65	26.24	1.58	21.89	2.70			4.10	2.54	3.67	16.45	10.62	9.61	2.00	1.25	9.21	0.89	3.50	1.00		2.76		0.68	1.85
	PE3	157	**180.38**	29.60	**176.76**	49.75	38.17	**249.09**		**791.84**	**246.01**	**384.19**	88.78	10.13	89.62	41.24	**171.65**	31.08	16.88	2.64	28.58	79.50	128.80	88.66	108.76	22.27	1.77	7.07	17.02	3.86	17.01	6.38	2.88	6.01	1.40	8.59	2.60
	PE4	129	125.82	5.02	4.51	0.14	2.32	41.50		120.52	78.52	**227.03**	126.76	1.20	59.03	**179.74**	**330.27**	156.40	38.51	2.53	71.63	63.72	44.02	47.91	132.60	13.47	0.94	1.70	4.25	3.00	8.11	1.58	6.08	1.03		0.29	2.85
	PE5	227	**254.78**	53.27	8.82	108.32	3.76	**359.47**	21.28	**420.67**	78.10	126.03	64.48	10.44	96.28	**581.01**	**223.78**	**251.82**	166.67	2.65	**175.42**	**179.37**	148.80	102.23	**172.22**	32.89	1.42	3.44	2.88	13.52	25.41	3.04	2.30	3.53	12.58	6.31	5.01
	PE6	167	43.94	**181.40**	30.70	**441.61**	19.34	**3004.09**	114.99	144.81	81.51	138.62	48.38	142.91	149.77	11.51	90.80	44.11	5.36	0.46	10.22	42.57	22.58	21.72	26.59	7.15	11.13	2.37	3.24	5.52	33.20	0.63	0.64	53.66	16.54	16.94	2.00
	PE7	120	71.04	71.88	**359.24**	**311.71**	147.28	**402.84**	8.93	166.47	**290.84**	**304.50**	14.46	29.72	145.34	2.35	45.67	7.16	5.78	0.33	5.58	3.50	16.78	26.77	13.26	0.83	15.57	2.81	8.87	0.42	6.69	0.20	1.73	17.01	1.78	0.68	0.96
	PE8	122	57.41	14.74	2.53	0.70	3.20	50.36	2.13	148.35	49.50	105.37	30.56	14.82	93.18	33.57	**201.37**	56.17	36.75	2.97	34.94	51.69	86.56	24.99	98.22	8.71	7.32	4.28	7.11	13.48	44.75	3.64	2.78	12.44	1.65	2.31	5.45
	PE9	21		0.20	4.33		0.33	10.57		0.17	1.00	11.40	0.33	1.00	1.50	0.50	4.00	0.50		0.67					3.00	1.00								0.50			
	PE10	33	8.06	0.50		0.85	0.33	13.51		0.60	2.73	33.07	6.04	1.27	2.02	4.10	12.28	3.99	2.97	0.76	5.49	0.08	1.00	0.75	2.30	3.25			4.65		2.53	2.45	0.33	7.66		1.00	6.42
Social Humanities	SH1	2						1.83	0.33					0.83			1.00																				
	SH2	4	0.40			1.67		0.33						0.20			1.00																	0.40			
	SH3	14	2.10	1.56	1.18	3.19	1.63	**19.96**	8.28		0.10	3.49		4.10	**42.54**	0.33	2.08	3.56		0.50				0.50		0.38			0.25		0.24		0.83	0.20			
	SH4	47	0.33	2.78	0.21	1.58		**93.89**	10.67		1.17	3.82	6.34	2.39	**63.99**	1.35	**22.26**	9.95		1.32	0.25						0.29							1.95	0.83	0.25	1.38
	SH5	4			0.07			3.86						0.07		0.50	1.00	0.50																			
	SH6	8				0.67		**11.37**	0.33	0.50	0.50	0.38	1.25	0.83		1.75	**23.78**				0.50	0.50			0.45						0.20						
Synergy Grant		12	1.86		2.90	0.75	1.87	29.57	0.50	2.82	5.57	**67.53**	25.05	2.04	12.31	6.45	**59.12**	52.16	1.00	2.00	2.76	0.58	0.06	1.72	3.83						0.33			0.88			0.33
Total Patent Applications by field*			731.07	367.10	622.83	913.90	221.70	4708.13	172.25	1598.84	1050.50	1721.83	1344.03	229.20	1291.09	1879.69	7186.88	5251.62	287.98	191.99	429.03	390.66	387.57	331.37	623.16	112.56	43.60	23.21	54.96	43.17	233.67	15.11	16.02	100.51	37.48	45.19	31.07

Note: * A patent application can cite publications linked to more than one ERC-funded project (e.g. to a LS1 and to a PE2 project). For such reason, in this Table the sum of values related to patent applications from the various technology field is higher than the total number of unique patent applications (reported in the final row of the table).

Table 21 - Distribution of citing patents by ERC sub-sector (row) and WIPO Technology Fields (column), H2020 grants (2014-2016) – Nr of Citing Patent Applications in cell

Sector	Subsector	N Projects	Electrical Engineering								Instruments					Chemistry											Mechanical Engineering								Other Fields		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Life Sciences	LS1	26						0.87		0.67	0.67	1.90	21.67		0.36	7.31	**75.54**	**40.44**	1.38	3.44				0.33	0.14						3.29						
	LS2	33						9.32				1.29	12.25		3.26	2.66	**108.81**	**38.89**		1.01	2.99		0.67		1.48	0.11	0.06		0.11		1.11						
	LS3	18		0.38				2.50			5.37	3.56	7.14		0.86	0.83	25.63	11.26						1.14	1.00	0.33					1.00						
	LS4	36						3.50		1.00		0.25	12.43		8.17	9.74	**60.01**	**63.97**		1.32			0.97						0.20	0.60	2.61		0.08		0.15		
	LS5	24		0.22		1.00		7.97		0.25	2.67	2.03	3.53		4.00	4.39	16.15	33.05					0.25		0.50												
	LS6	35			0.36	3.11		3.82				2.88	13.10		1.67	10.65	**295.55**	**115.08**		4.40	1.07			0.18	2.00						5.14						
	LS7	76	4.08	2.67	0.50			24.64		0.50	1.13	25.09	29.47	0.49	**41.52**	20.60	**117.86**	**120.98**		2.05	1.65				3.44	2.58			3.80		3.41		0.14		1.40		
	LS8	11						2.00								1.20	15.31	9.87		4.57	1.10										1.95						
	LS9	28	1.87					1.50				2.42	2.91		1.83	2.35	**60.04**	4.97	1.63	5.39	2.33			0.84	2.42	2.08					0.42						
Physical Engineering	PE2	31	4.41	2.73	5.67	0.40	1.17	18.15		1.60	**31.78**	18.66	0.50		0.44						0.50		1.60	3.15	1.74	2.00			0.50								
	PE3	32	6.77	0.84	2.18		3.07	**24.73**		**36.22**	11.57	**61.07**	8.48	2.41	1.36	2.09	21.67	4.52	0.96		4.75	1.60	3.50	7.09	5.79	1.00		0.25		0.15	0.88			0.72			0.33
	PE4	30	1.25				0.50	2.33		5.76	5.13	11.28	4.08		0.33	9.92	6.50	2.00	2.39		11.98	**43.78**	1.17	5.18	**49.89**	14.33					0.20						
	PE5	58	14.63		0.09	0.29		4.86	0.14	9.67	1.19	8.47	8.58		0.13	**23.64**	14.67	**29.24**	14.27		15.26	8.21	12.59	5.77	14.88	1.60		0.79	1.79	2.35	7.54	0.25				0.13	
	PE6	54	1.26	9.72	2.30	**40.10**	0.71	**410.77**	4.53	2.63	4.07	**22.18**	1.03	**29.34**	5.44	0.47	0.17	0.49			0.82	1.92		0.81	0.58		1.42		0.67	0.37	1.10	1.00		10.14	1.17	2.79	
	PE7	52	4.46	6.73	**28.62**	17.47	1.89	**72.40**	0.40	**36.17**	**68.84**	**40.50**	1.27	3.88	19.07		3.00		0.67		0.75	0.25	1.50	3.14			0.90				12.50			1.58			
	PE8	45	8.63	0.67		1.00		6.64		7.90	5.90	16.32	6.04	0.34	6.30	0.89	10.93	7.06	3.05		5.99	10.61	2.18	2.26	9.05	2.67	2.00	0.60	2.17	1.57	6.33	1.50			1.40		3.00
	PE9	3		0.25				0.20		1.80	0.75																										
	PE10	3									0.25	1.75									1.00																
Social Humanities	SH1	1															1.00																				
	SH3	2		0.33		0.33			0.33																	1.00											
	SH4	8						0.67					1.00		**4.00**	0.33	1.00	1.67																	0.33		
Total Patent Applications by field*			45.22	23.47	39.36	62.70	7.34	577.85	5.41	99.45	136.27	209.60	123.18	36.47	92.70	94.71	769.88	457.29	24.34	18.59	48.00	64.45	24.42	29.28	91.56	26.10	4.38	1.64	9.23	5.04	43.36	2.75	0.22	12.44	3.06	2.92	3.33

Note: * A patent application can cite publications linked to more than one ERC-funded project (e.g. to a LS1 and to a PE2 project). For such reason, in this Table the sum of values related to patent applications from the various technology field is higher than the total number of unique patent applications (reported in the final row of the table).

On the other hand, ERC Physical and Engineering projects show a broader impact, as they span multiple WIPO macro categories. In particular, the scientific fields showing a higher influence are Fundamental Constituents of Matter (PE2), Condensed Matter Physics (PE3), Physical and Analytical Chemical Sciences (PE4), Synthetic Chemistry and Materials (PE5), Computer Science ad Information (PE6), Systems and Communication Engineering (PE7), and Products and Processes Engineering (PE8). We observe a significant impact of PE8 on Biotechnologies, which signals that new manufacturing techniques are critical for the industry. While all general ERC grants can be considered leaning towards fundamental research, coherently with Cohen et al. (2002) the scientific fields highly cited by patents gather projects more directly linked to specific applications or the solution of specific societal challenges. Following this line of reasoning, Mathematics (PE1), Universe Science (PE9), and Earth System Science (PE10) show a lower level of direct technological influence. The one divergence from this pattern is the impact on patents observed for projects from the scientific field Fundamental Constituents of Matter (PE2), related for instance to the development of instruments technologies (optics and measurement).

The Social Sciences and Humanities ERC projects are much less related to any patenting activities, as we have previously seen in this report. The three matrices reaffirm this finding, but it is also interesting to notice the impact of The Human Mind and its Complexity (SH4) on computer technology (6), medical technology (13), biotechnology (15), and pharmaceuticals (16). Similarly, the projects developed on The Social World and its Diversity (SH3) had an impact on computer technology (6) and medical technology (13).[22]

In Table 22, we offer another attempt to represent the impact of ERC projects on the generation of new technologies—this time, by considering the general distribution of global patent applications by WIPO classes. The intent is better understand whether the technological specialisation patterns of citing patents mimic the more general specialisation patterns of overall patenting activity[23]. For each WIPO field, we calculate the ratio between the share of patent applications in that field citing a publication coming out of an ERC project and the share of total EPO applications in that field. The greater this number, the higher the impact of ERC-funded project citations for that specific class. The reason being that their contribution to the field is proportionally greater than their share in the general distribution of patent applications. In particular, a value of this ratio higher than 1 provides evidence of a positive specialisation in that specific technology field of patents.

[22] This evidence could be influenced by the relative number of projects funded by ERC in each category. While we cannot completely rule out this possibility, we did consider these values in additional analyses, together with the single class-normalised mean difference. The high level of variance does not seem to signal any impact of inter-class differences.

[23] In order to construct the denominator of such an indicator, we used the OECD patent by technology database as a source of data. For each WIPO technology field, we identified the share of global patent applications in that field out of total global patent applications. We referred to EPO patent applications and used the 2008-2019 period of analysis to construct this value, since more recent data were largely incomplete.

Table 22 - Specialisation Index of patents citing ERC projects' publications, for all grants in the sample, FP7 grants, H2020 grants (positive specialisation for values > 1)

Wipo Technological Sector	WIPO Technological Field	ERC	FP7	H2020
Electrical Engineering	1 - Electrical machinery, apparatus, energy	0.32	0.32	0.21
	2 - Audio-visual technology	0.39	0.39	0.29
	3 - Telecommunications	0.76	0.77	0.56
	4 - Digital communication	0.39	0.39	0.27
	5 - Basic communication processes	0.93	0.97	0.35
	6 - Computer technology	**2.23**	**2.19**	**2.58**
	7 - IT methods for management	0.41	0.42	0.12
	8 - Semiconductors	**2.26**	**2.30**	**1.64**
Instruments	9 - Optics	**1.39**	**1.38**	**1.83**
	10 - Measurement	**1.10**	**1.10**	**1.29**
	11 - Analysis of biological materials	**4.63**	**4.62**	**4.83**
	12 - Control	0.42	0.42	0.59
	13 - Medical technology	0.54	0.55	0.39
Chemistry	14 - Organic fine chemistry	**2.00**	**2.04**	**1.22**
	15 - Biotechnology	**7.39**	**7.43**	**8.34**
	16 - Pharmaceuticals	**3.70**	**3.74**	**3.55**
	17 - Macromolecular chemistry, polymers	0.43	0.42	0.36
	18 - Food chemistry	0.68	0.68	0.75
	19 - Basic materials chemistry	0.54	0.53	0.64
	20 - Materials, metallurgy	0.72	0.66	**1.10**
	21 - Surface technology, coating	0.80	0.81	0.51
	22 - Micro-structural and nanotechnology	**6.70**	**6.76**	**7.05**
	23 - Chemical engineering	0.89	0.84	**1.31**
	24 - Environmental technology	0.32	0.29	0.75
Mechanical Engineering	25 - Handling	0.05	0.05	0.05
	26 - Machine tools	0.03	0.03	0.02
	27 - Engines, pumps, turbines	0.05	0.05	0.08
	28 - Textile and paper machines	0.08	0.08	0.10
	29 - Other special machines	0.24	0.22	0.37
	30 - Thermal processes and apparatus	0.03	0.03	0.05
	31 - Mechanical elements	0.02	0.02	0.00
	32 - Transport	0.06	0.06	0.07
Other Fields	33 - Furniture, games	0.06	0.06	0.05
	34 - Other consumer goods	0.06	0.07	0.04
	35 - Civil engineering	0.03	0.03	0.04

Four fields show values of the specialisation index greater than 3, for both the whole sample and the programme-specific projects: Biotechnology, Micro-Structural and Nano Technology, Analysis of Biological Materials, and Pharmaceuticals. Five more fields—Semiconductors, Computer Technology, Organic Fine Chemistry, Optics, and Measurement—show values greater than one for both the whole sample and the programme-specific projects. Finally, Chemical Engineering and Materials and Metallurgy show values greater than one, but only for H2020 projects. Such results confirm the findings of previous studies of patent-to-paper citations (Ahmadpoor and Jones, 2017), which are summarised in Section 3 of this report. In short, the number of citations to scientific literature in patents varies dramatically across technology fields, but is much more pronounced in emerging fields and those at the forefront of the science-to-technology frontier[24].

[24] In additional analyses not reported here we combined Figure 18 with Figure 21, presenting the same matrix combining ERC classes and WIPO classes in rows and columns but reporting in each cell the specialization index. The cells with values greater than 1 were considered in order to highlight a relative specialization in a WIPO technology class, as compared to the overall patent applications distribution. This set of analyses confirm the broader spectrum of influence of projects developed within Physical Engineering classes and the higher concentration of Life Science ones. All ERC subsectors show an index higher than one in at least one WIPO class. Biotechnology, Analysis of Biological Materials as

We repeated all these analyses on a different sample of patent applications linked to ERC grants: namely, those self-reported by all ERC beneficiaries in the period of observation as direct output of the different projects, as shown in Table 23. There are 1,550 declared patent applications directly generated by the grantees. It is worth noting that the results illustrated in Table 23 generally confirm all reported patterns, in spite of the different size of the two samples examined. Thus, we are more assured of these patterns' robustness.

Micro-Structural and Nano Technology WIPO classes show consistently large values of the specialisation index distributed across different ERC panels.

Table 23 - Distribution of self-reported patents by ERC sub-sector (row) and WIPO Technology Fields (column) – Nr of self-reported Patent Applications in cell

Sector	Subsector	N Projects	Electrical Engineering								Instruments					Chemistry											Mechanical Engineering								Other Fields		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Life Sciences	LS1	34	1.50					3.05			1.83	2.12	5.84		0.25	2.87	68.17	26.29				0.50		0.33	0.50						0.75						
	LS2	36	0.75					1.33				0.20	4.09			1.17	37.17	7.63		1.91					0.61						0.33		0.80				
	LS3	24						0.50			1.50	1.00	3.20		0.53	1.39	11.81	11.56	0.40	0.33	0.17				0.40						0.20						
	LS4	39						1.30				0.20	11.48		2.10	0.92	27.59	35.21							0.50		0.25				0.45						
	LS5	30		0.33	0.50			1.63			1.75	0.25	2.25	1.00	13.88	0.83	18.93	16.15					0.67		0.17						2.33		1.00		0.33		
	LS6	38										0.50	7.04			2.17	22.53	23.89		0.37					0.50												
	LS7	85						4.53			2.00	19.29	4.55		28.06	7.38	55.40	57.13		1.27	1.30			1.26	2.29	1.00			1.00		0.80					1.73	
	LS8	5						1.00				2.33	0.33				2.17								1.17												
	LS9	28	0.65									2.08	1.70		3.20	1.54	38.48	6.32		0.50	0.65	1.15		0.25	0.29	1.18				1.00				1.00			
Physical Engineering	PE2	25	1.00	0.75	0.50	2.00	0.33	3.13		0.38	17.13	8.37			1.50	1.00								2.67	2.25	1.00			0.50					0.50			
	PE3	44	5.84	2.05	1.27	1.00	2.64	10.78	0.14	21.43	5.00	10.04	2.70	1.43	1.00		4.25				0.42	1.20		1.09	4.77	0.42			0.45		1.60	0.09	1.25	1.00		0.14	
	PE4	50	7.96				0.20	3.36		6.13	2.25	23.76	5.82		2.25	7.19	16.12	3.50	1.83	0.17	3.70	4.66	1.00	5.47	16.21	1.25					1.17						
	PE5	88	14.04	0.17		0.40		2.82	0.20	13.20	6.67	4.50	5.01		5.41	40.68	15.99	15.68	6.99	2.50	7.56	18.95	12.73	3.38	10.43	3.62		0.58		2.42	1.75		1.33				
	PE6	37	2.58	3.30	1.78	32.39	1.75	43.17		1.57	3.09	8.70	0.60	4.07	1.28	0.38	0.70					0.25		2.05									1.00	1.67		0.67	
	PE7	61	15.80	3.94	12.78	3.55	16.30	33.48		26.26	53.00	45.04	0.83	0.10	7.51		1.00	0.17			1.00	0.14	0.20	5.66	5.25		0.50	0.14			0.50				0.50	0.33	
	PE8	58	8.46	0.83			1.00	3.83		14.88	1.67	9.58	1.66		6.23	1.56	5.29	6.24	0.82		3.23	7.41	4.96	4.43	11.05	1.75		0.60	1.29	1.25	1.63	1.05	1.33	1.25		3.50	3.23
	PE10	7	1.00									1.75	0.50				0.75				2.00			0.25	0.75												
Social Humanities	SH3	1						0.50	0.25																										0.25		
	SH4	3						4.00																													
	SH6	1															1.00																				
Synergy Grant		4						0.14		0.22	1.00	2.99			0.50		1.50	0.50						0.14	2.00												
Total Patent Applications by field*			56.90	11.38	16.84	39.34	22.23	115.54	0.59	84.06	93.89	136.71	56.58	5.60	72.70	68.75	326.35	208.60	10.05	7.05	20.03	34.27	19.55	26.39	57.93	10.22	0.75	1.33	3.24	4.67	11.52	1.14	6.72	5.42	1.08	6.37	3.23

Note: * A patent application can be reported by more than one ERC-funded project (e.g. to a LS1 and to a PE2 project). For such reason, in this Table the sum of values related to patent applications from the various technology field is higher than the total number of unique patent applications (reported in the final row of the table).

6.3 A focus on patents related to Climate Change Mitigation and Adaptation Technologies

We now provide a more focused analysis of patented inventions inspired by ERC research projects—those related to the pool of climate change mitigation and adaptation technologies (CCMATs). To perform these analyses, we adopted the classification schemes developed by the EPO (as described in section 6.1). Using this scheme, we identified 3,026 citing patent applications related to CCMAT technologies, corresponding to 8.77% of total patents citing publications of ERC projects in the period 2008-2016. In order to provide a benchmark to compare this value, we can consider the share of patent grants at the EPO in environment-related technologies, which represents around 12% of total EPO patent grants (period 2015-2018, OECD data - https://stats.oecd.org/). Accounting for the limitations of this simple comparison (based on different time periods and different classification approaches), we do not observe a specific specialisation of ERC-linked patents in this area.

If we analyse such results at the project level, we notice that such 3,026 CCMAT patents cite publications coming from 731 projects (out of 6,671 projects included in our sample). Therefore, around 11% of ERC projects included in our time window generated scientific outcomes cited in patented technologies related to climate change mitigation and adaptation[25]. As shown in Table 24, a deeper analysis of the 3,026 CCMAT-related patent applications by technology sectors shows the prevalence of two macro-sectors: 37.93% of patents belong to the sub-field Y02E “Reduction of greenhouse gas (GHG) emissions, related to energy generation, transmission or distribution” and 34.57% of them belong to the sub-field Y02A “Technologies for Adaptation to Climate Change”. Another well-represented sub-sector is Y02P “Climate change mitigation technologies in the production or processing of goods” with a percentage of 15.25%. These subfields are also the most represented in the case of self-reported patents (where we found 136 patents in this area out of 1,550, corresponding to a share of 8.8%), with percentages of 43.14% for Y02E (higher than the percentage of citing patents), 34.07% for Y02A and 15.68% for Y02P.

[25] On a similar vein, out of the 6,671 projects in the dataset, the total number of projects cited by patent applications is 2,977. And out of those 2,977 projects, there are 731 cited by patents on climate change mitigation and adaptation technologies. Therefore, around 24.5% of the projects cited by subsequent patents are cited by patented inventions related to climate change mitigation and adaptation technologies.

Table 24 - Distribution of patent applications citing ERC publications related to CCMAT technologies, by sub-group (fractional counting; all projects in the sample)

CPC Classification		Citing Patents		Declared Patent Applications	
		Number of Patent Applications	Share (%) of Patent Applications	Number of Patent Applications	Share (%) of Patent Applications
Y02E	Reduction of greenhouse gas (GHG) emissions, related to energy generation, transmission or distribution	1,147.75	37.93%	58.67	43.14%
Y02A	Technologies for Adaptation to Climate Change	1,046.00	34.57%	46.33	34.07%
Y02P	Climate change mitigation technologies in the production or processing of goods	461.58	15.25%	21.33	15.68%
Y02D	Climate change mitigation technologies in information and communication technologies (ICT)	110	3.64%	1	0.74%
Y02C	Capture, storage, sequestration or disposal of greenhouse gases (GHG)	75.17	2.48%	0.5	0.37%
Y02T	Climate change mitigation technologies related to transportation	74.42	2.46%	1	0.74%
Y02B	Climate change mitigation technologies related to buildings	44.75	1.48%	3.84	2.82%
Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. Smart grids	43.83	1.45%	0.33	0.24%
Y02W	Climate change mitigation technologies related to wastewater treatment or waste management	22.5	0.74%	3	2.21%
Total		***3,026.00***	***100.00%***	***136***	***100.00%***

Table 25 reports the distribution of CCMAT-related patent applications across the different ERC sectors, based on the category of the ERC project cited (through its publications) by the patents. The concentration of CCMT-related patents is particularly pronounced in the group of patents linked to PE projects (in this group of patents, around 10.62% of patents can be categorised as related to sustainable technologies, according to the scheme developed by the EPO). A lower percentage characterises the group of patents linked to LS (in this group of patents, around 6.75% of patents are in the sustainable domain). In the SSH sector, only 0.77% of cases are related to CCMAT patents. The analysis of self-reported patents shows similar percentages: 10.73% for PE and 6.3% for LS. For projects related to the SSH sector, we were not able to identify self-declared patents related to CCMATs.

Table 25 - Distribution of ERC projects and related citing patent applications by CCMATs (by ERC sector of the project)

Sector	Total projects (a)	Total Climate Change-related patent applications, by sector (b)	Number of projects cited by at least one Climate Change-related patent application (c)	% projects cited by at least one Climate Change-related patent application (c/a)
PE	3,077	1,942	371	12.06%
LS	2,372	1,119	354	14.92%
SSH	1,203	3	3	0.25%
Synergy Grant	19	17	3	15.79%
Total*	6,671	3,026	731	10.96%

*Note: * A patent application can be reported by more than one ERC-funded project (e.g. to a LS and to a PE project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique patent application (reported in the final row of the table).*

6.4 A focus on patents related to the Digital Transformation

We now closely examine the set of patented inventions linked to ERC-grants and related to technologies relevant for the Digital Transformation, using the classification of Fourth industrial Revolution patents developed by the EPO and described in section 6.1.

We first provide the results for patent applications relying on publications from ERC projects. We identified 7,583 patent applications in this group related to 4IR technologies, corresponding to 22.30% of total patents citing publications of ERC projects. This share is rather stable if we compare patents linked to FP7 projects (where the share of 4IR related patents out of total patents is around 21.9%) and those linked to H2020 projects (where the share is 22.58%). In order to have a reference value to better interpret this evidence, we considered the statistics provided by the last edition of the EPO report on 4IR technologies (EPO, 2020). We observed that 4IR patents in 2018 accounted for more than 11% of all patenting activity worldwide, thanks to a substantial growing trend over the previous years. A comparison with such a number (with the caution required by the consideration of two different time periods) seems to suggest a valuable specialisation of ERC-linked patents in this area.

If we analyse such results at the project level, we notice that such 7,583 4IR patents cite publications coming from 1,071 projects (out of 6,671 projects included in our sample). Therefore, around 16% of ERC projects in the dataset are cited at least by one 4IR related patent application.

If we breakdown these 7,583 4IR patent applications by technology sectors, we find that 40.21% of them are included in the Enabling Technologies macro-sector, 32.98% in the Core Technology sector, and 26.81% in the Application Domains sector. According to the EPO report, Application Domain sector is the dominant one in the general distribution of all global 4IR patents. However, in the case of 4IR patents inspired by ERC-research we notice a prevalence of technologies linked to the basic building blocks (Enabling Technologies and Core Technologies).

We replicated the analyses for the group of self-reported patents and identified a total of 239 patent applications related to Fourth Industrial Revolution technologies, corresponding to 15.45% of total patents directly produced as a result of ERC projects. This share is a little below the corresponding share found for citing patents, showing a more pronounced specialisation in the latter case. A deeper look of 4IR self-reported patents by sector shows that 38% of them refer to the Enabling Technologies Domains, 33% on Core Technology Domains, and 29% on Application Domains.

Table 26 - Distribution of patent applications citing ERC-related publications in the 4IR technology fields, by sub-group (fractional counting; all projects in the sample)

4IR Technology Sector	4IR Technology Field	Patent citing publications		Self-reported patents	
		Number of Patent Applications	% Patent Applications	Number of Patent Applications	% Patent Applications
Core Technologies	Connectivity	880.7	11.61%	40.17	16.81%
	It hardware	930.66	12.27%	35.5	14.85%
	Software	689.84	9.10%	15	6.28%
Enabling Technologies	Core AI	59.81	0.79%	0	0.00%
	Data management	1826.77	24.09%	28.42	11.89%
	Data security	399.36	5.27%	21.17	8.86%
	Geo positioning	95.06	1.25%	6.67	2.79%
	Power supply	26.51	0.35%	1	0.42%
	Safety	10.59	0.14%	0	0.00%
	Three-dimensional support system	258.01	3.40%	5.83	2.44%
	User interfaces	372.99	4.92%	16.33	6.83%
Application Domains	Agriculture	21.84	0.29%	1.75	0.73%
	Consumer goods	211.56	2.79%	4.92	2.06%
	Healthcare	1055.45	13.92%	39.42	16.49%
	Home	36.9	0.49%	1.58	0.66%
	Industrial	86.67	1.14%	4.25	1.78%
	Infrastructure	41.52	0.55%	0.58	0.24%
	Services	414.12	5.46%	14.92	6.24%
	Vehicles	164.64	2.17%	1.5	0.63%
Total		**7583**	**100.00%**	**239**	**100.00%**

If we look at the finer classification of 18 different 4IR technological fields, we notice that the field including the highest share of 4IR-related patent applications linked to ERC-funded research is "Data Management" (corresponding to 24.09% of all 4IR citing patents), followed by "Healthcare" (13.92%), "IT Hardware" (12.27%) and "Connectivity" (11.61%). The Data Management field (including inventions related to diagnostic and analytical systems for massive data, prediction and forecasting techniques, monitoring functions, planning and control systems) is thus the prevalent one, involving around one quarter of all 4IR patent applications citing ERC-related scientific publications. It is interesting to notice a significant share of 4IR patents in the application domain of Healthcare (encompassing, e.g., technologies related to intelligent healthcare systems, robotic surgery, smart diagnosis), which signals a specialisation in this area (which is ranked ninth out of 18 4IR technology fields in terms of the number of 4IR patents in 2018, according to the

EPO report on global 4IR patents). The corresponding values for the group of self-reported patents are slightly different: the most represented category is “Connectivity” (16.81%), followed by “Healthcare” (16.49%), “IT Hardware” (14.85%), and “Data Management” (11.89%). In this group of 4IR self-reported patents, we observe a significant presence of inventions related to applications in the healthcare domain.

Table 27 reports the distribution of 4IR citing patent applications across the different ERC sectors, based on the category of the ERC project cited (through its publications) by the patents. As expected, the concentration of 4IR patents is particularly pronounced in the group of patents linked to PE projects (where nearly 34% of patents can be categorised as related to the Fourth Industrial Revolution, according to the scheme developed by the EPO), and more limited in the group of patents linked to LS (where only around 7% of patents are in the 4IR domain). It should also be noted that patents citing ERC projects from the SSH sector are related to 4IR in 68.63% of cases, highlighting the relevance of this technological domain for patents citing ERC publications. A similar distribution is observed in the case of self-reported patents, where we observe that 5 out 6 self-reported patent applications resulting from SSH projects are related to 4IR technologies.

If we disaggregate further, we can see that patents citing publications from ERC projects in the LS sector are quite concentrated in the 4IR sub-sector «Application Domains»: 60.86%, mostly in the Healthcare category. Patents citing ERC projects in the PE sector are more dispersed across technological sectors: 42.64% of them are related to Enabling Technologies, 36.35% to Core Technologies, and 21.01% to Application Domains.

Table 27 - Distribution of ERC projects and related citing patent applications by 4IR technologies (by ERC sector of the project)

Sector	Total projects (a)	Total 4IR-related patent applications, by sector (b)	Number of projects cited by at least one 4IR-related patent application (c)	% projects cited by at least one 4IR-related patent application (c/a)
PE	3,077	6,326	679	22.07%
LS	2,372	1,148	328	13.83%
SSH	1,203	267	58	4.82%
Synergy Grant	19	44	6	31.58%
Total*	6,671	7,583	1,071	16.05%

*Note: * A patent application can be reported by more than one ERC-funded project (e.g. to a LS and to a PE project). For such reason, in this Table the sum of values related to projects from the various domains is higher than the total number of unique patent application (reported in the final row of the table).*

6.5 A focus on projects highly cited by subsequent patents

As a final step in these analyses of patents by technology domain, we identified a set of ERC FP7 and H2020 projects whose publications were highly cited in patent documents. To perform these analyses, we ranked projects by number of citations received from patent applications (via NPL). For each highly cited project, we used secondary sources (Cordis, project websites and dedicated web-searches) to collect brief background information on the research topics of interest, the main technological developments, and the valorisation outcomes.

This type of analysis is helpful for providing qualitative information that contextualises highly influential projects and their technological achievements. It can also be useful for identifying some areas where the knowledge stemming from ERC projects was particularly valuable in inspiring technological developments. Table 28 below reports the first 10 projects by patent citations received by the respective articles. The table shows that some projects significantly influenced subsequent inventions, thanks to hundreds of received citations. The projects reported in this table came from the FP7 programme—as expected, due to the longer time span available for them to influence subsequent technological developments. As the table shows, 5 projects are from the LS sector and 5 are from the PE sector.

Table 28 - ERC projects receiving the highest number of citations from subsequent patents (via publications cited in the NPL; all projects in the sample)

Project	Programme	Acronym	Project Title	Sector	N of patent applications citing project's articles
228180	FP7	VisRec	Visual Recognition	PE	740
249845	FP7	TARGETING GENETHERAPY	Towards Safe and Effective Hematopoietic Stem Cell Gene Therapy: Targeting Integration to Genomic Safe Harbors and Exploiting Endogenous microRNA to Regulate Transgene Expression	LS	498
261063	FP7	BRAINCELL	Charting the landscape of brain development by large-scale single-cell transcriptomics and phylogenetic lineage reconstruction	LS	396
232814	FP7	StemCellMark	LGR receptors mark adult stem cells in multiple mammalian tissues	LS	346
207354	FP7	Graphene	Physics and Applications of Graphene	PE	322
279881	FP7	HYPER	Hybrid Photovoltaic Energy Relays	PE	302
247404	FP7	MESOLIGHT	Mesoscopic Junctions for Light Energy Harvesting and Conversion	PE	289
232989	FP7	GrowthControl	Dissecting the transcriptional mechanisms controlling growth during normal development and cancer	LS	285
208813	FP7	ChemBioMech	Exploring mechanism in chemical biology by high-throughput approaches	PE	282
320339	FP7	ImmunoDeath	Immunogenic cell death in anticancer therapy	LS	280

In order to complement such evidence, we also provide information for the set of highly cited H2020 projects during the period 2014-2016 (Table 29).

Table 29 - H2020 ERC projects receiving the highest number of citations from subsequent patents (via publications cited in the NPL; H2020 projects 2014-2016)

Project_id	Programme	Acronym	Project_title	Sector	Number of patent applications citing project's articles
639707	H2020	REMEMBER	Adaptive immunity in prokaryotes: how bacteria do not forgive and do not forget their enemies	LS	198
677268	H2020	nextDART	Next-generation Detection of Antigen Responsive T-cells	LS	116
671093	H2020	SynCatMatch	MATching zeolite SYNthesis with CATalytic activity	PE	111
647769	H2020	RSM	Rich, Structured Models for Scene Recovery, Understanding and Interaction	PE	109
669598	H2020	SynDiv	A nanophysics approach to synthetic cell division	PE	88
638009	H2020	IDIU	Integrated and Detailed Image Understanding	PE	68
670603	H2020	ZAUBERKUGEL	Fulfilling Paul Ehrlich's Dream: therapeutics with activity on demand	LS	57
681760	H2020	TransModal	Translating from Multiple Modalities into Text	PE	57
640079	H2020	QPE	Quantum Photonic Engineering	PE	53
680916	H2020	HYPHEN	HYPHEN: Hybrid Photonic Engines for Massive Cloud Connectivity	PE	41

Table 30 – Publications included in our dataset receiving the highest number of citations from subsequent patents

Publication title	Author(s)	Year of Publication	Journal	ERC Project(s)	Programme	N of patents citing	N of Scopus citations (July 2022)
Very Deep Convolutional Networks for Large-Scale Image Recognition	Simonyan, Karena and Zisserman, Andrew	2015	3rd International Conference on Learning Representations, ICLR 2015	228180	FP7	469	15,947
Counting absolute numbers of molecules using unique molecular identifiers	Kivioja, Teemua, et al.	2012	Nature Methods	232989 / 261063	FP7	260	547
SLIC superpixels compared to state-of-the-art superpixel methods	Achanta, Radhakrishna, et al.	2012	IEEE Transactions on Pattern Analysis and Machine Intelligence	247170	FP7	217	6,148
Efficient coupling of light to graphene plasmons by compressing surface polaritons with tapered bulk materials	Nikitin, A. Yu., et al.	2014	Nano Letters	258461	FP7	196	77
An updated evolutionary classification of CRISPR–Cas systems	Makarova, Kira S., et al.	2015	Nature Review Microbiology	639707	H2020	189	1,320
A foundation for universal T-cell based immunotherapy: T-cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR.	Torikai, Hiroki, et al.	2012	Blood	249845	FP7	163	335
An unbiased genome-wide analysis of zinc-finger nuclease specificity	Gabriel, Richard, et al.	2011	Nature Biotechnology	249845	FP7	157	412
A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins	Corti, Davide, et al.	2011	Science	250348	FP7	148	869
Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease	Anders, Carolin, et al.	2014	Nature	337284	FP7	137	718
Fiji: an open-source platform for biological-image analysis	Schindelin, Johannes, et al.	2012	Nature Methods	260746	FP7	133	26,660

Table 30 reports the publications included in our dataset that received the highest number of citations by subsequent patent applications along with their respective original projects. Interestingly, in the vast majority of cases, such publications also had an outstanding impact in the scientific domain, as highlighted by the high values of citations received by scientific articles (based on Scoupus data, as of July 2022). In order to confirm this suggestive evidence in a more systematic and robust way, in Annex 4 at the end of the report we perform a set of regression analyses based on the final sample of ERC funded FP7 and H2020 projects. In such models, we analyze the effect exerted by the scientific excellence of an ERC project (as captured by the number of Scopus scientific citations received by its publications) and its likelihood of being cited by subsequent patents (in the NPL). The results of such regression models confirm the existence of a positive and statistically significant relationship between such variables, after controlling for a set of influential factors at the project level. This evidence thus supports previous research findings highlighting the existence of a positive relationship between scientific impact (as measured by the number of citations received in peer-reviewed publications) and technological impact (as captured by patent-to-paper citations).

The analyses of the technology domains for highly cited FP7 and H2020 projects, coupled with the analyses reported in the previous three sections, helped us to identify innovative technological developments where the intellectual influence exerted by ERC-funded research has been valuable. For instance, by simply looking at the Cordis description of the ERC projects reported in Tables 28 and 29 (representing the FP7 and H2020 projects with the highest number of citations from subsequent patents), it is possible to identify influential contributions in these dynamic and impactful technology fields:

Image recognition technologies: in the context of computer vision, encompassing the ability of software to identify objects, places, people, writing and actions in images, thanks to deep learning-based approaches.

Graphene: the unique properties of this material open up a large number of transformative applications in transport, medicine, electronics, energy, defence, and desalination.

Solar cell technologies: related to the development of innovative semiconducting materials and next-generation solar cells, able to accelerate the achievement of maximum solar-to-electricity efficiency at low cost.

Application of microRNAs: they have rapidly emerged as promising targets for the development of novel anticancer therapeutics.

Immunotherapy treatments: using a person's own immune system to fight cancer.

Stem cell technologies: applicable in a variety of contexts (regenerative medicine, test of new drugs for safety and effectiveness, understanding how diseases occur).

This evidence can be adopted in a second step of the study on major innovations inspired by ERC-funded research results to better identify a set of technologies that form the core of a small number of selected impactful innovations (existing or forthcoming). In this regard, the previous list of technologies is certainly not exhaustive, as it only refers to a small set of highly influential ERC projects (based on patents citing their publications). Thus, the list should be integrated with other data sources, such as the evidence coming from academic startups and spinoffs inspired by the results of ERC-funded projects or from the outcomes of ERC Proof-of-Concept projects resulting in successful valorisation trajectories (Munari and Toschi, 2021).

In conclusion, we report below a selected number of short cases, taken from a selection of the highly influential projects mentioned above, in order to in order to have a more vivid, qualitative picture of their technological contributions. In order to derive a general overview of all ERC sectors, we present below a selected number of projects in the LS, PE and SSH domains that received the highest number of patent citations (through publications). We focus first on FP7 projects and then on H2020 projects.

PE Sector (FP7)

Project name and acronym	***VisRec - Visual Recognition*** (228180), FP7 project
Sector	**Physical Engineering**
PI and Host Institution	Andrew Zisserman, University of Oxford
Patent applications citing the project (via NPL):	740 (citing 47 projects' scientific publications)
Technologies	This project developed a visual system that is able to learn, recognise and retrieve quickly and accurately thousands of visual categories, including objects, scenes, human actions and activities, called by them as "Visual Google" (for images and videos).
Project Contribution	Progress has been made on a number of fronts including: (i) learning visual models on-the-fly to retrieve semantic entities in large-scale image and video collections starting from a text query; (ii) automatic identification of flower species and sculptures; (iii) methods and models for detecting and localizing object categories in images - in particular, reducing the level of supervision that is required when training such models; and (iv) deep learning methods for recognising object categories, text, and human actions and inter-actions (such as handshakes) in images and videos.
Valorisation	The results of this research project supported the development of several software (https://www.robots.ox.ac.uk/~vgg/projects/visrec/). The PI and his colleagues also funded a company called "Vision Factory" that was acquired by DeepMind (part of the Alphabet Group) in 2014. The spin-off company is based on the technologies developed by the PI and his research group on Visual Recognition.

Project name and acronym	***Graphene - Physics and Applications of Graphene*** (207354), FP7 project
Sector	**Physical Engineering**
PI and Host Institution	Konstantin Novoselov, University of Manchester
Patent applications citing the project (via NPL):	322 (citing 43 projects' scientific publications)
Technologies	The project is a pioneer in the development of applications to graphene, exploiting the PI's lead in the emerging research area (the PI was awarded with the Nobel Prize in 2010). The project covered three main directions, exploring most the exciting features of graphene.
Project Contribution	According to the project's final report, graphene has become a gold mine for searching for new phenomena, offering numerous applications from smart materials to future electronics. The project investigated graphene ballistic field effect transistors, single electron transistors, gas sensors and other electronic devices.
Valorisation	The graphene presented among the years several applications from electronics to composite materials. This new material has great conductivity and strength properties, and has been used by several sectors and companies. There are also 4 self-reported patents linked to this project.

Project name and acronym	***HYPER - Hybrid Photovoltaic Energy Relays*** (279881), FP7 project
Sector	**Physical Engineering**
PI and Host Institution	Henry James Snaith, University of Oxford
Patent applications the project (via NPL):	302 (citing 20 projects' scientific publications)
Technologies	The project developed Photovoltaic (PVC) solar cells that promise to be a major contributor to the future energy supply.
Project Contribution	The landmark development has been the discovery of solid-state organic-inorganic metal halide perovskite solar cells, which deliver over 20% power conversion efficiency and are set to rival the performance of crystalline silicon, but with the advantage of very inexpensive materials, low temperature processing and versatile solution or vapour phase manufacture. Beyond the application, there has been a plethora of new physics fundamental understanding of materials and hybrid interfaces developed within the project. The next 10 years of optoelectronics research into perovskites and the associated compounds promises to be rich, productive and fulfilling.
Valorisation	The project generated two startups: One founded by the PI, called Helio Display Materials that creates photoluminescent and electroluminescent perovskite-based materials for the display industry that use significantly less power. Recently, it raised an investment around \$4.75 million. Another startup founded by the PI and his team is Oxford Photovoltaics, which develops and commercialises a perovskite-based solar cell technology. There are also 3 self-reported patents linked to this project.

LS Sector (FP7)

Project name and acronym	***TARGETINGGENETHERAPY - Towards Safe and Effective Hematopoietic Stem Cell Gene Therapy: Targeting Integration to Genomic Safe Harbors and Exploiting Endogenous microRNA to Regulate Transgene Expression*** (249845), FP7 project
Sector	**Life Sciences**
PI and Host Institution	Luigi Naldini, Universita Vita-Salute San Raffaele
Patent applications citing the project (via NPL):	498 (citing 30 projects' scientific publications)
Technologies	The project developed a more effective Hematopoietic Stem Cell (HSC) Gene Therapy that regulates transgene expression by exploiting cellular microRNAs. It also targets integration at predetermined sites of the genome by forcing homologous recombination with designer Zinc finger nucleases.
Project Contribution	Starting from an initial identification of miRNA that show preferential activity in Hematopoietic Stem and Progenitor Cells (HSPC), the project designed regulated vector cassettes containing an optimised combination of target sequences complementary to such miRNA, thereby suppressing unwanted transgene expression in HSPC. They describe promising results on the treatment of tumors. The project demonstrated that gene-edited human HSC are able to repopulate transplanted mice and give rise to functional immune cells. Based on these promising results, the team is optimising/scaling-up the HSPC gene editing procedure in order to develop a clinically applicable protocol.
Valorisation	The research team, including the PI, founded a clinical-stage biotechnology company to treat cancer with a technology platform that allows the direct delivery of immunotherapeutic payloads within the tumor microenvironment. The technology developed by the PI and improved by the project is nowadays a reference to the development of tools in biomedical research. This project also generated 1 self-reported patent.

Project name and acronym	***BRAINCELL - Charting the landscape of brain development by large-scale single-cell transcriptomics and phylogenetic lineage reconstruction*** (261063), FP7 project
Sector	**Life Sciences**
PI and Host Institution	Sten Linnarsson, Karolinska Institute
Patent applications the project (via NPL):	396 (citing 16 projects' scientific publications)
Technologies	They determined cellular lineage using somatic mutations and cellular identity ("cell type") using large-scale, single-cell transcriptomics.
Project Contribution	They developed synthetic constructs in transgenic mice, currently based on a lentiviral barcoding strategy. This approach will not reveal complete lineage trees, but will label sub-lineages stochastically. They also built a high-quality dataset of more than 25,000 single cells, covering mouse brain and some other tissues (notably hair follicles, a key model of adult stem cells).
Valorisation	The PI's lab focuses on developing technologies for extremely sensitive and accurate detection of RNA in single cells. Based on the knowledge and technology developed by the PI and his research group, Karolinska Institute (in partnership with KTH Royal Institute of Technology, Stockholm University and Uppsala University) founded a Laboratory in 2010 called SciLifeLab, which is focused on the advancement of molecular biosciences. There are also 3 self-reported patents linked to this project.

Project name and acronym	***StemCellMark - LGR receptors mark adult stem cells in multiple mammalian tissues*** (232814), FP7 project
Sector	**Life Sciences**
PI and Host Institution	Johannes Carolus Clevers, Royal Netherlands Academy of Arts and Sciences
Patent applications citing the project (via NPL):	346 (citing 27 projects' scientific publications)
Technologies	They identified stem cells of multiple internal organs using the Lgr5 marker previously characterised in their lab, as well as its family members.
Project Contribution	They identified a number of novel stem cells, including those of the stomach, the liver, the pancreas and the sebaceous glands. Furthermore, the project developed technology that allows for the in-vitro expansion if these stem cells, against the dogma that non-transformed cells cannot be propagated in a culture without transformation. The stem cells identified in this project and the methods to expand them indefinitely in vitro holds great promise for the study of basic aspects of adult stem cell biology, as well as patient-specific drug development platforms. The technology also offers a novel and genetically stable source of cells for regenerative and/or gene therapy.
Valorisation	The technology is currently implemented through the generation of large biobanks consisting of cultures of patient-specific disease samples such as tumour organoids or organoids from cystic fibrosis patients. The biobank platform provides the first test of patient-specific samples for their response to conventional or new therapies. The PI founded three startups in connection with this project. The first one is Hubrecht Organoid Technology, a global leader in the field of adult stem cell-derived organoids, according to their website. The second one is Surrozen, which provides a therapeutic to treat diseases characterised by tissue injury using targeted regenerative antibodies to repair a broad range of tissues and restore organs damaged by serious disease. The third one is Xilis, which develops a new therapeutic technology called MicroOrganoSphere™ to help cancer patients and accelerate drug discovery. There are also 2 self-reported patents linked to this project.

SSH Sector (FP7)

Project name and acronym	***SYNPROC - Synchronous Linguistic and Visual Processing*** (203427), FP7 project
Sector	**Social Science and Humanities**
PI and Host Institution	Frank Keller, The University of Edinburgh
Patent applications citing the project (via NPL):	45 (citing 5 projects' scientific publications)
Technologies	Linguistic input often occurs synchronously with visual input, e.g., in everyday activities such as attending a lecture or following directions on a map. The visual context constrains the interpretation of the linguistic input, and vice versa, making processing more efficient and less ambiguous.
Project Contribution	The project developed an experimental research programme that investigated key features of synchronous processing by tracking participants' eye movements when they view a naturalistic scene and listen to a speech stimulus at the same time. The aim is to understand synchronous processing better by studying the interaction of saliency and ambiguity, as well as the role of incrementality, object context, and task factors.
Valorisation	The results fed a series of computational models that predict the eye-movement patterns that humans exhibit when they view a scene and listen to speech at the same time.

Project name and acronym	***GMI - Genetics of Mental Illness*** (230374), FP7 project
Sector	**Social Science and Humanities**
PI and Host Institution	Dorret Boomsma, Stichting VU
Patent applications citing the project (via NPL):	40 (citing 21 projects' scientific publications)
Technologies	Genome-wide association (GWA) scans is a technique that can detect the genetic architecture of some mental disorders (i.e., migraines and depression).
Project Contribution	With a close focus on Attention Problems (AP) and anxious-depression (AD), this project carried out large genetic studies of these traits in children, adolescents and adults. The project addresses three interrelated topics: 1. Neuropsychiatric disorders and cognition; 2. Depression, anxiety, substance use, abuse and dependence; 3. Depression, migraine and cardiovascular risk. It also developed models for the genetic analysis of these complex traits, especially gene-environment interaction and genome-wide association models. The goal is to discover which genes influence the risk for mental disorder and co-morbid biomedical traits, to identify the causal variants, and to explore their interaction with environmental risk factors.
Valorisation	The results showed that the genetics of common mental disorders indicate a high degree of polygenic inheritance, explaining the heritability of these disorders; a high degree of genetic pleiotropy, explaining the mutual comorbidity and the comorbidity with somatic disease; and a high degree of genetic stability across the lifespan, explaining the persistence of mental disorders.

LS Sector (H2020)

Project name and acronym	**REMEMBER** - ***Adaptive immunity in prokaryotes: how Bacteria do not forgive and do not forget their enemies*** (639707), H2020 project
PI and Host Institution	Stan Brouns, TU Delft
Patent applications the project (via NPL):	198 (citing 5 projects' scientific publications)
Technologies	CRISPR is a technology that can be used to edit genes. The essence of CRISPR is finding a specific bit of DNA inside a cell and altering that piece of DNA.
Project Contribution	This project determines the mechanism of the enigmatic process of primed memory formation against heavily mutated invaders. Using a combination of genetic, biochemical and structural approaches, including state-of-the-art single molecule imaging of CRISPR immunity in living Escherichia coli cells, the project proves that perfectly matching and degenerate targets are differentially recognised, and trigger either target DNA degradation or priming. Moreover, the project tested whether degenerate priming is a universal phenomenon among different CRISPR-Cas types. Thus, degenerate priming will impair the use of viruses as therapeutic agents to treat antibiotic-resistant bacterial infections. The project screens for organic molecules that inhibit the formation of CRISPR resistance. These molecules can be co-administered with viruses to potentiate treatments. A better understanding of immune systems as stages for the ongoing evolutionary battle between viruses and bacteria could create opportunities to use viruses as an alternatives to antibiotics.
Valorisation	The PI has founded a national open-source bacteriophage biobank (www.fagenbank.nl) that can serve as a resource for phage therapy research.

Project name and acronym	**nextDART - *Next-generation Detection of Antigen Responsive T-cells*** (677268), H2020 project
PI and Host Institution	Sine Reker Hadrup, Technical University of Denmark
Patent applications the project (via NPL):	116 (citing 7 projects' scientific publications)
Technologies	The project developed a new technology based on multimerised peptide-major histocompatibility complex I (MHC I) reagents that allow the detection of >1000 different T-cell specificities with high sensitivity in small biological samples.
Project Contribution	The current ability to map T-cell reactivity to certain molecular patterns poorly matches the huge diversity of T-cell recognition in humans (about 107). Current state-of-the-art T-cell detection enables the detection of 45 different T-cell specificities. Therefore, a comprehensive analysis of T-cell recognition against intruding pathogens, auto-immune attacked tissues or cancer is virtually impossible. The project developed a novel technology based on multimerised peptide-major histocompatibility complex I (MHC I) reagents that allows for an in-depth understanding of T-cell recognition from a structural perspective. It can be used to evaluate clinical efficacy and safety profiles of T-cell receptors for clinical use.
Valorisation	The PI founded a company called Tetramer Shop, in partnership with other researchers. This company develops drug technology aims to MHC tetramers to make the testing and monitoring of antigen-specific T-cells.

PE sector (H2020)

Project name and acronym	***Rich, Structured Models for Scene Recovery, Understanding and Interaction* (University of Heidelberg)**
PI and Host Institution	University of Heidelberg
Patent applications the project (via NPL):	106 (citing 14 projects' scientific publications)
Technologies	Computer vision has gained considerable momentum in recent years, with the goal of bringing computer vision from the lab into real life.
Project Contribution	The project designs novel models, learning and inference techniques to make computer vision real and accurate. It has proposed the Rich Scene Model (RSM), which is one joint statistical, structured model of many physical and semantic scene aspects that can take full advantage of the synergy effect between all its components.
Valorisation	The PI developed more than 30 patents.

Project name and acronym	***Matching Zeolite Synthesis with Catalytic Activity***
PI and Host Institution	**Spanish National Research Council**
Patent applications the project (via NPL):	76 (citing 18 projects' scientific publications)
Technologies	Zeolites are solid, porous catalysts that have wide-ranging industrial applications for gas absorption, separation and catalysis. Solid catalysts are key components in many industrial processes, offering advantages such as reuse and ease of recovery. This research programme aims to develop a new concept and methodology for the synthesis of zeolite catalysts. While a relatively large number of zeolites have been synthesised, selecting a zeolite as a catalyst for a particular reaction still involves a large element of trial and error.
Project Contribution	The project developed a new synthesis methodology for zeolite that creates pores and cavities in the resulting zeolite that approach a "molecular recognition" pattern to catalyze the desired reaction. The approach focuses on the study of the reactions' transition states, as the literature accepts that the most efficient catalysts are those that lower the transition state energy in the reaction, thereby boosting the catalytic activity and efficiency.
Valorisation	The PI has founded the Institute of Chemical Technology (ICT), which has filed more than 150 patent applications, from which 80 derived from R&D agreements with industry.

7. Conclusions and recommendations

Fundamental science and research at the frontier of knowledge are essential to many facets of improving society: from individual health and well-being, to balanced development capable of reconciling growth and natural resources, to the generation of new ideas and models able to capture the evolution of values and norms. New ideas and discoveries are inherently risky and characterised by low appropriability—two conditions that tend to discourage private investments and require public involvement. The European Research Council is the main institution and instrument that Europe uses to pursue this goal.

In the past twenty years, there has been rising attention on the accountability and productivity of public investments—with fundamental science and research being no exception. Granted, the characteristics of these fields pose several challenges. First, the notion of impact can be complex, its measurement subject to different perspectives and interests. Second, indirect effects might be more relevant than direct ones, but these are inherently more difficult to anticipate and, therefore, evaluate. Third, the nonlinear development of innovation clearly highlights the relevance of serendipitous paths of knowledge from science to society, but this challenges the causal models that are typically used in impact studies.

This study was developed for the European Research Council for two purposes: (1) to contribute to the current debate on measuring how public funds allocated to fundamental science and research impact society; and (2) to produce a unique perspective on projects funded in the 7th framework and in H2020 programs. Following consolidated approaches in the literature on innovation economics, we developed two different indicators of knowledge development: scientific publications and patents. The simple intuition is to measure if, how and to what extent the new discoveries presented by ERC grantees in scientific publications inspired new technologies described in new patents. We used the publications cited in patents applications to trace this influence.

Our main conclusion is straightforward: Based on the data collected and all different analyses performed, the citations arising from ERC-funded projects exert considerable influence upon patentable technology. More than 40% of ERC grants generate research that is subsequently cited by patents, which is significantly higher than the share of grants directly generating self-reported patents (around 10%). These results provide specific evidence of the indirect effect of research on technological development, which would not have emerged from an analysis based solely on the specific outputs reported by each grantee.

Coherently with the few other studies that have presented similar analyses in different countries and programmes, we observed significant variation across fields regarding the influence exerted upon technology development. We show that this variation applies to both the research fields and disciplines of the grantees, as well as on the technological classification of patents based on WIPO classes and more recent aggregations focusing on sustainable transition and industrial applications associated with the fourth industrial revolution.

This variation is mainly associated with the number and diversity of technological fields. We find general evidence that patent citations to ERC-funded research often flow across technological fields, consistent with the idea of widespread diffusion of frontier research results. This flow is more widespread for projects funded in the Engineering and Science fields, while Life Science projects tend to be more concentrated.

In short, the percentage of funded projects and the extent of their presence in several technological fields represent two clear ways of connecting science to innovation. That said, we also tried to refine our depiction of the relevance of this influence: namely, by showing that, in many fields, ERC-funded research is cited significantly more than research developed through other programmes.

Not surprisingly, it takes time for science to influence technology and promote innovation. Our results clearly show that the H2020 projects are comparable in relative terms to FP7 ones, but are fewer in absolute terms due to their more recent deployment. Similarly, it might not be easy to blur the boundaries between institutions and roles. Specifically, we documented that, in the case of self-reported patents, universities and research institutions hold more patents citing ERC-funded research than firms or individuals, whereas in the case of citing patents in around 40% of cases universities or research institutes result as applicant. The geographic distribution of influence is another interesting aspect that could be further explored based on our collected data. Patents' citations of scientific publications implies that knowledge flows through dedicated communication channels. A closer look at the physical distance between the patent holders and the grantees (i.e., based on their working addresses) could offer additional evidence. If closeness would emerge as dominant, it might suggest the role of other forms of dissemination and localised networking opportunities.

We also tried to contribute to the debate on whether there are differences in fundamental research projects that impact science vs. technological development. While the literature on star scientists tends to focus on individuals, our findings emerged from a project-based design focused on a specific set of policy measures targeting excellent science. Most of these programmes target junior- and middle-career researchers, and we are therefore able to observe those who will become star scientists. Our evidence, based on regression analyses linking the number of patent citations received by ERC projects to the number of scientific citations they received from the literature, supports previous findings highlighting that those who are highly visible and recognised by scientific communities, based on the number of citations received in peer reviewed publications, are also highly cited by patents.

Our study confirms that the grants-publications-patents approach is a promising way to investigate the technological influence of publicly funded work. Nonetheless, our experience with it underscores some areas of improvement. First, as already documented by patent-citation studies, we need to better understand how citations are used by different types of patent owners, how citations change over time, or how they are affected by reforms of IPR institutions. Second, in the era of Big Data, the quality and reporting of collected information are often taken for granted, but they still require great attention and care. In this specific project, we noticed an opportunity to improve the quality of information on patents self-reported by ERC grantees, due to the multiplicity of formats used by beneficiaries in their communication.

While WIPO categories are widely used by many institutions to perform the difficult task of clustering similar technologies, more sophisticated data analysis techniques based on natural language processing could improve the identification of the most influential fields. One opportunity could be given by the patent landscape analysis, adopting topic modeling techniques based on systematic keywords search. Alternative options to be explored could be topic modelling based on latent Dirichlet allocation (Blei, 2012), where topics are represented as clusters of similar words expressing the hidden semantic structure of a set of documents, or keyword extraction (Moretti et al., 2015), with relevant multiword expressions being ranked by frequency to represent the main semantic domains in a group of texts.

Of course, patents are only one way through which we can trace the impact of new scientific discoveries. Further studies could integrate the grant-publication-patent flow with data on new companies that are founded to commercially exploit new knowledge. Although with different methodological challenges, it could be explored evidence on products and processes linked to the patents observed. Another opportunity is exploring the practice of patent reassignments of both cited and self-reported patents in the sample to dig deeper into the impact of the original research funded.

Importantly, we adopted a descriptive approach with no intention of drawing strong causal conclusions. The project was funded to perform a set of innovative and necessary activities to build the dataset. By combining multiple sources and new methodologies, we were able to generate a robust data platform that can be used to develop more sophisticated analyses—perhaps accounting for potential endogeneity or unobservable variables biases. One possible approach that would be coherent with the data structure would be a matched-paired analysis based on a control group of patented technologies that are homogenous by field and year of development. There may also be value in collecting and maintaining information on the ERC applicants whose projects were not selected by the evaluation panels, as such information could be used to develop counterfactual evaluations.

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583-587.

Angelucci, S., Hurtado-Albir, F. J., & Volpe, A. (2018). Supporting global initiatives on climate change: The EPO's “Y02-Y04S” tagging scheme. *World Patent Information*, *54*, S85-S92.

Azoulay, P.; Graff Zivin, J. S.; Li, D.; Sampat, B. N. (2015). 'Public R&D investments and private-sector patenting: evidence from NIH funding rules'. NBER: Working Paper.

Azoulay, P.; Graff Zivin, J. S.; Li, D.; Sampat, B. N.. (2019). 'Public R&D investments and private-sector patenting: evidence from NIH funding rules'. The Review of Economic Studies, 86(1), 117-152.

Blei, D. (2012) Probabilistic Topic Models, in Communications of the ACM, vol. 55, n. 4, April 2012, pp. 77–84.

Bryan, K. A., Ozcan, Y., & Sampat, B. (2020). In-text patent citations: A user's guide. *Research Policy*, *49*(4), 103946.

Callaert, J.; Pellens, M.; Van Looy, B. (2014) ‘Sources of inspiration? Making sense of scientific references in patents'. Scientometrics, 98 (3), 1617-1629.

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

Cohen, W. M.; Nelson, R. R.; Walsh, J. P. (2002) 'Links and impacts: the influence of public research on industrial R&D'. Management Science, 48 (1), 1-23.

de Moya-Anegon, F., Lopez-Illescas, C., Guerrero-Bote, V., & Moed, H. F. (2020). The citation impact of social sciences and humanities upon patentable technology. *Scientometrics*, *125*(2), 1665-1687.

Du, J.; Li, P.; Guo, Q.; Tang, X. (2019) 'Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis'. Journal of Infometrics, 13 (1), 132-148.

European Patent Office. (2017). Patents and the Fourth Industrial Revolution.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. Strategic management journal, *25*(8-9), 909-928.

Fleming, L., Greene, H., Li, G., Marx, M., Yao, D. (2019). 'Government-funded research increasingly fuels innovation'. Science, 364(6446), 1139-1141.

Grassano, N.; Rotolo, D.; Hutton, J.; Lang, F.; Hopkins, M.M. (2017) 'Funding data from publication acknowledgments: Coverage, uses, and limitations'. Journal of the Association for Information Science and Technology, 68 (4), 999-1017.

Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993). 'Geographic localization of knowledge spillovers as evidenced by patent citations'. The Quarterly Journal of Economics, 108(3), 577-598.

Jefferson, O. A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U., Bilder, G. (2018). 'Mapping the global influence of published research on industry and innovation'. Nature Biotechnology, 36(1), 31-39.

Li, D., Azoulay, P., Sampat, B. N. (2017). 'The applied value of public investments in biomedical research'. Science, 356(6333), 78-81.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal, *41*(9), 1572-1594.

Mazzoleni, R., & Nelson, R. R. (2007). Public research institutions and economic catch-up. Research policy, *36*(10), 1512-1528.

Moretti, G., Sprugnoli, R. & Tonelli, S. (2015). Digging in the dirt: Extracting keyphrases from texts with kd. In Proceedings of CLiC it 198 (2015).

Munari, F. & Toschi, L. (2021). The impact of public funding on science valorisation: an analysis of the ERC Proof-of-Concept Programme, Research Policy, 50(6).

Narin, F., Hamilton, K.S., Olivastro, D., 1997. The increasing linkage between U.S. technology and public science. Research Policy, 26 (3), 317–330. https://doi.org/10.1016/ s0048-7333(97)00013-9. OECD, 2009

Poege, F.; Harhoff, D.; Gaessler, F.; Baruffaldi, S. (2019). 'Science quality and the value of inventions'. Science Advances, 5(12), 1-8.

Rai, A.K., Sampat, B. N. (2012). 'Accountability in patenting of federally funded research'. Nature biotechnology, 30(10), 953-956.

Tijssen, R.J.W., Buter, R.K., Van Leeuwen, T.N. (2000). 'Technological relevance of science: Validation and analysis of citation linkages between patents and research papers'. Scientometrics, 47, 389–412.

van Raan, A. F. (2017). Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research. Journal of Data and Information Science, *2*(1), 13-50.

van Raan, A. (2019). Measuring science: Basic principles and application of advanced bibliometrics. In Springer handbook of science and technology indicators (pp. 237-280). Springer, Cham.

Veugelers, R., Wang, J. (2019). 'Scientific novelty and technological impact'. Research Policy, 48 (6), 1362-1372.

Winnink, J. J., Tijssen, R. J., & van Raan, A. F. (2013). The discovery of introns: Analysis of the science-technology interface. In*Translational Twists and Turns: Science as a Socio-economic Endeavor. Proceedings of the 18 th International Conference on Science and Technology Indicators* (pp. 427-438).

Annex1: Description of methods and data

In these sections we describe in more detail the methods used to collect, process, and aggregate the data used in the study, concerning grants, publications and patents.

1. ERC projects identification

As a first step, a comprehensive list of projects awarded with an ERC grant within FP7 and H2020 programs was compiled. This list relies on data provided by ERCEA. The projects identified were 4,556 within the FP7 frame, and 7,832 for H2020. The types of grants considered are ERC Starting Grants (ERC-StG), ERC Consolidator Grants (ERC-CoG), ERC Advanced Grants (ERC-AdG), ERC Proof of Concept Grants (ERC-PoC), ERC Synergy Grants (ERC-SyG).. For each ERC project key information was retrieved, i.e. *project-ID*, *acronym*, *title*, *call*, *type* of ERC, *start and end date*, *sector and subsector*, *PI name*, *PI surname*, *coordinator, participants*.

Table A 1 - Number of ERC projects divided by H2020 and FP7 and types of grants considered

	ERC-StG	ERC-CoG	ERC–PoC	ERC-AdG	ERC–SyG	Tot. Projects
FP7	2,771	2,553	1,125	1,584	99	4,556
H2020	2,332	313	178	1,709	24	7,832

2. Publications dataset

This step aims to create a dataset with the information about all the publications that have been developed with the support of one or more ERC projects identified in the previous section. Particularly relevant is information about *title*, *DOI*, *authors* and *year of publication*, as we will use them in the successive steps. To create this ERC grant-publication dataset, we used five different and (partially) independent datasets of publications and organised them into a unique dataset. In this dataset, the overlapping publications that were found in the different datasets were merged, but the relative metadata from all the different sources were preserved. The different sources of publications utilised are: (i) Cordis[26] (i.e. European Commission's primary public portal used to disseminate the research results produced from EU-funded research), (ii) Scopus[27] (i.e. Elsevier's database of publications), (iii) a dataset of publications provided by ERCEA, (iv)

[26] https://cordis.europa.eu/

[27] https://www.scopus.com/

publications databases retrieved from the European portal for open data[28] and (v) an ERC publication dataset built by some of the authors of this report in 2018 for the EPO ARP research project.

Using all these different datasets can introduce noise in the data. However, the rationale behind the choice of using multiple data sources is that retrieving all the information available maximises the likelihood of having a match with the publications' information from the Non-Patent-literature table of the Patstat database, after which an in-deep quality control is performed. The following sections will describe in detail how the ERC-publications dataset has been created.

2.1. Publications dataset from Cordis

On Cordis website, each ERC project has a dedicated page and a section "Results" containing bibliographic references to the publications that have been supported financially by the project. In Cordis publications are intended in a broad sense, not only papers but also books' chapters, conferences proceedings, thesis. To create our dataset, the section "Results" of each FP7 and H2020 project of interest has been downloaded via web-scraping and the relevant information for each publication was extracted.

2.2. Publications dataset from Scopus

Scopus is an abstracting and indexing database of scientific publications being published in virtually all scholarly journals of any significance in the world. Scopus analyses the funding sources cited in the acknowledgement's sections of scientific publications and inserts them in a dedicated metadata field. Scopus provides APIs to gather these types of information. Here, we enquired Scopus via API for retrieving all the publications citing in the acknowledgment section one of the following "European Research Council" or "ERC" or "Horizon 2020 Framework Programme" or "Seventh Framework Programme" or "H2020" or "FP7". Together with one of these keywords, one ERC grant number from the ERC-funded projects (respectively FP7 or H2020) had to be present in the acknowledgment section. Noteworthy to notice, the information available for each publication retrieved from Scopus is highly homogeneous and detailed.

2.3. Publications datasets from ERCEA, previous research and OpenEU

Three additional datasets were used. The first dataset has been provided by ERCEA, which provided us with a dataset of publications for H2020 and one for FP7. Additional datasets for both H2020 and FP7 have been retrieved from the website managed by the Publication Office of the European Union (official portal for distributing the EU open data), where is possible to download a dataset of publications for ERC FP7 and

[28] https://data.europa.eu/en

H2020. The last dataset used comes from a study conducted by some of the authors of this report in 2018 for the EPO ARP research project. This dataset was built using Cordis webpage and Scopus database.

2.4 Duplicates removal

Before merging the different datasets into one, we removed the publications duplicates within each dataset. Sometime duplicates can be noise (errors), and in this case only one of the entries is maintained. Other times the same publication is supported by more than one ERC project and thus has a different entry for each. In this case, the entries are considered as one, but a reference is kept to all the projects that funded it.

Duplicates were removed by matching the strings of title and DOI between all the publications. If DOI and/or title of two or more publications matched, they were merged into one. To improve matching performances, before matching, titles and DOIs were normalised. Titles by turning text to lowercase and removing accents and non-alphanumeric characters. DOI were stripped of unnecessary parts (e.g. the "https://doi.org/", sometime preceding the actual DOI) and checked to be properly formatted in order to avoid incomplete DOIs. Strings not corresponding to DOIs are not considered as they consist of information filled in the wrong field (e.g. page numbers). In case of multiple matches, a match on DOI has priority over a match on the title.

After duplicates removal, final datasets contained the following publications shown in Table A 2below.

Table A 2 - Number of publications for each dataset after duplicates removal, before the organisation of the datasets into one.

Source	FP7	H2020
Cordis-dataset	119,225	73,285
Scopus-dataset	100,419	73,206
Ercea-dataset	106,423	72,324
OpenEu-dataset	137,688	66,163
prevRes-dataset	116,025	40,615

2.5 Final dataset creation

The aim of this step is to converge the different sources into a single final dataset. We aim to identify the unique publications of the datasets. When the same publication is contained in multiple sources we merged it into a single entry, keeping all the metadata available from the different datasets to maximise the information available. To this end, we first merged two datasets: if a match between two publications of the two datasets occurred, the two publications were considered the same entry and merged. On this merged dataset, one of the remaining datasets was also matched with the same procedure and so on through all the five datasets. In the resulting dataset each publication can be the result of merging across two or more datasets or contained in only one dataset.

While merging, we considered a match when two publications have the same title and/or DOI. To increase the quality of the matches, reducing the number of false negatives, the title is normalised by turning text to lowercase and removing accents and non-alphanumeric characters (e.g. apostrophes and dashes). DOIs were stripped of unnecessary parts (e.g. the “https://doi.org/” preceding the actual DOI in some instances) and checked to be properly formatted to avoid incomplete DOIs. Noteworthy, if two publications matched across datasets in terms of DOI, but titles were different, the two publications were matched as one entry, but the two versions of the titles were kept. This case is particularly relevant for publications containing for example chemical formulas, that very often are having encoding problems and vary in their plain text version.

Although this step can introduce some noise, the reasoning behind this is to maximise the probability to find a publication in Non-Patent-Literature. While a stricter check will be implemented during the phase of matching with the non-patent-literature for the moment our goal is to retrieve and keep all the information.

2.6 FP7 publications’ dataset cleaning and description

After merging the datasets as described in the previous section, the final dataset for FP7 consists of 198,751 publications. The dataset from our previous research has a function of “sanity check”, rather than adding new information, as the dataset was a combination of Cordis and Scopus. About 98% of the publications from the 2018s’ research is present also in one or more of the other datasets, while the 2% missing publications have been removed from the relative cordis page or Scopus with respect to 2018 when previous research was conducted. This confirms that the current research of publications is at least as good as the one of 2018.

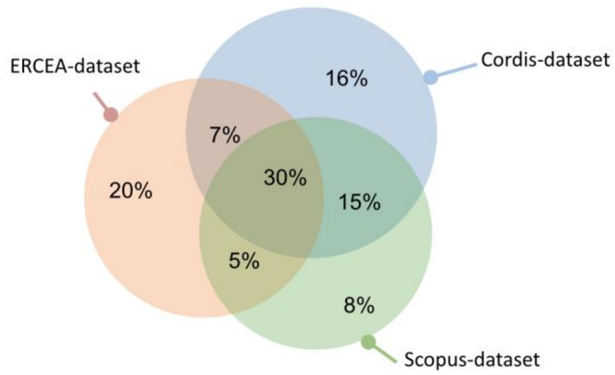
Not considering the previous research dataset (as it is contained in the other), about 51,000 publications (~26%) were found in all the other 4 datasets, 23,500 (~12%) in three datasets, 60,000 (~30%) in two datasets, while 61,000 (~30%) in one dataset. Given the fact that the different datasets are from different sources, we consider publications found in more than one dataset as reliable. Regarding the 61,000 publications found in only one dataset, we considered the part found only in Cordis (~1,644) or given to us by ERCEA (~2,200) as reliable. Regarding the part of publications found only in Scopus (~6,000), we performed a further control to ensure pertinence of the publication to ERC. For each publication, we controlled whether the relative PI is among the authors, resulting in 3,061 publications. Only these were included in the final dataset.

2.7 H2020 publications' dataset cleaning and description

Regarding H2020, after merging the datasets as described in the previous section, the final dataset consists of 97,875 publications. 98.5% of previous research results present in other datasets thus confirming also for H2020 that the current dataset is as good as the previous one. Not considering results from previous research for the motivations already explained in the previous section, about 45,000 publications (46%) are contained in all the 4 other datasets. The dataset from openEU is basically contained in Cordis. Cordis dataset, however, has around 7,000 more publications, possibly due to a more recent update. A residual part is contained only in the dataset from Cordis (~700) or ERCEA (~2,200) and were considered reliable.

However, about 21,000 publications have been retrieved only in Scopus. As for the FP7, in order to avoid introducing publications not traceable to an ERC project, we kept only those where the PI of the project was among the authors of the publication, that is 14,923 publications.

Figure A 1 - Overlap between the three main different sources of publications considered for Fp7 (left) and H2020 (right).



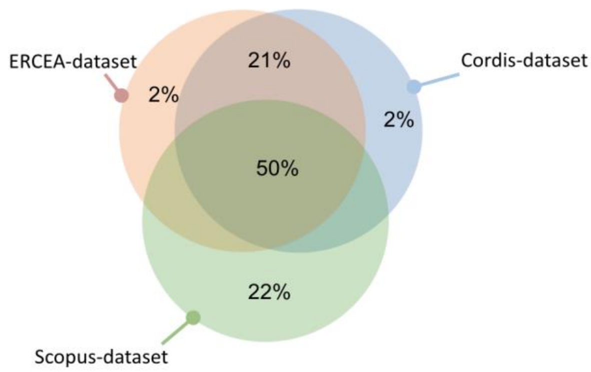


Figure A 1 shows the overlap of data among Cordis, Scopus and the ERCEA dataset. For H2020 we observe a considerable overlap between the ERCEA-dataset and the Cordis-dataset. However, Scopus differs and contains a substantial number of publications not retrieved in the other datasets. This can be attributed to the fact that Scopus identifies ERC-publications through the acknowledgment section, and this section has become mandatory only starting from more recent ERC-projects. It is likely that many ERC-publications have missed to acknowledge the ERC during FP7, as it was not mandatory.

2.8 Retrieve of additional information from Scopus

To uniform the information available for each publication, we retrieved from Scopus using the DOI, when available, the Scopus metadata of the publications of the final dataset that were not in the Scopus dataset. Between publications retrieved in Scopus via the acknowledgments and publications retrieved in Scopus via

DOIt, 144,364 publications for FP7 and 80,918 publications for H2020 were identified in Scopus. For each publication found in Scopus, we downloaded also the distribution of citations over the years.

3. Patents-dataset creation

The following steps describe how ERC publications found in previous steps, are linked to patents. We used Patstat , the database released by the European Patent Office (EPO). Patstat provides access to patents' data from more than 40 patent authorities worldwide and is updated every six months. In the current research the 2021-Autumn edition is used.

We consider a link between a patent and ERC publication, when the publication is cited in the NPL section of the patents. Patstat is organised so that each citation referring to a scientific publication, i.e. "Non-Patent-Literature" (NPL), is reported in a table (the NPL table tls214). In the version of Patstat used for the current research, this table contains more than 32 million entries that have been matched with the ERC publication dataset so that any entry in common between the two was retrieved. The following steps describes how this procedure was conducted and the number of patents retrieved that were citing an ERC publication.

3.1 Linking ERC-publications with Non-Patent-Literature (NPL) in Patstat

Once we have defined the list of relevant publications produced with the support of an ERC grant, we matched them with the references figuring as Non-Patent Literature (NPL) in Patstat (table *tls214*). The NPL table of Patstat contains the scientific publications cited in each patent.

Metadata for NPL publications in table *tls214* of Patstat can be found in two configurations. There are structured entries, with a separate field for every information (e.g. *title*, *authors*, *DOI* or *year*). Those consist of about 9 million entries. Conversely, unstructured data, where all the information about the publication is merged into a single free-text field, is the majority (about 22 million). To extract the output of ERC-projects related to patents we matched the 31M NPL references in Patstat with the publications linked to FP7 and H2020 projects from the previous step.

The match is based on three criteria: having the same first author (surname), same title and same DOI. To note that the FP7 and H2020 publications' datasets were built so that for each publication it was possible having multiple versions of the title and DOI as previously explained. In these cases, the match is done by considering every variant of the title/DOI, so to maximise the chances of finding a match. To note also that author surname is available only for the publications that were found in Scopus, as this data-source provides a specific field for first author surname, while the metadata relative to first author of other datasets are often provided in a free-unstructured form and thus difficult to isolate.

As first step, the match on the DOIs was performed. Afterward, before performing the match for author and title, all the data were normalised (in both our dataset and the NPL table entries). All the text is set to lowercase and accents on letters and non-letter characters (e.g. apostrophes and dashes) are removed. This allows to overcome issues when matching surnames differing from each other for the presence of an accent or titles having a formula in them that was written slightly differently.

For NPL instances with structured metadata, we match the information from ERC publications only with the related Patstat field (e.g., table *TLS214_NPL_PUBLN*, fields: *npl_author*, *npl_title*, *npl_doi*). Here we consider only publications resulting from exact matches, where the entire surname, title or DOI is the same. Matches due to substring are not considered (e.g. *"organic compounds"* does not match with *"macromolecular organic compounds")*. When dealing with unstructured metadata, we match *author*, *title* and *DOI* with all the information available regardless of the position. In this case matches can occur also with substrings.

Each pair of publications (one by an ERC grant and one in NPL-Patstat) can match with one or more of the searched strings: *title*, *DOI* and *author*. A higher number of hits means a more confident match (e.g. if a publication matched for *title*, *DOI* and *author* is the best result we can obtain). In total we have 7 possible combinations, e.g. match only with *DOI*, only with *title*, only with *author*, with *DOI* and *author*, with *title* and *author*, with *DOI* and *title*, with *author*, *title* and *DOI*.

A priori, we discarded matches based only on first author surnames, given the very high number of results produced and the low significance due to instances of homonymy, while we considered as valid the matches with two or more hits (a match with two features is very unlikely to occur by chance). The publications that matched only for *title* or for *DOI* were manually inspected. Often a match occurred only for title (or DOI) occurred in presence of short titles (or short DOIs), e.g. a publication in the ERC dataset titled "The Alzheimer disease" was counted as substring of many publications in the NPL table. This is an example of the possible wrong matches that were controlled manually and discarded.

Table A 3 - Total number of matches after application of our matching algorithm and process of cleaning between our dataset of publication and NPL-table from Patstat

	FP7	**H2020**
title, doi, author	9,880	955
doi	81	8
doi, author	547	55
title	2,759	368
Title, author	22,970	2,015
Title, doi	664	120
Total	36,901	3,521

As shown in Table A 3, after the manual cleaning, we found 36,901 matches between the Patstat NPL-table and the papers from FP7 projects. In particular, 13,218 publications from our dataset generated a total of 57,256 citations (in 37,885 different applications for patents, see next section). This means that a paper in NPL can be cited by more than one patents and also that one patent can cite more than one publication. After the manual cleaning, we found 3,521 matches between Patstat and the papers from H2020 projects. In particular, 2,127 different publications from our dataset of publications generated a total of 5,296 citations (in 4,357 different applications for patents, see next section).

3.2 From Non-Patent-Literature to patents

Once identified the entries in the NPL-table of Patstat related to ERC-grants, via queries to Patstat database is possible to identify the patents' applications that cited the NPL-publication, passing from table NPL table tls214 to patents' applications table *tls201* of the patents' databasae (for the architecture of the database and more details about table contents see Patstat database description, footnote 20 of this report). It is worth to note the following: the unique identification ID of the application of a patent is usually linked to more versions of the same patent that is usually presented several times (the pat_publn_id, Patstat table *tls211*). In the following we consider each unique application, as it is the central unique identifier around which the Patstat database is built. Thus, starting from the NPL table of Patstat (table tls214), we queried the Patstat database and for each publication found in the previous match, we identified the application id (appln_id, table tls201) of the patent to which it is linked (passing through Patstat tables tls212, and tls211). We identified 40,346 unique patents, 37,885 of them are related to FP7 publications and 4,357 to H2020 ones. (with 1,896 patents related to both FP7 and H2020 projects).

Considered the information found on our sources, we used in our final dataset just publications whose bibliographic information was available in SCOPUS and that started after the beginning of the project, and their related patent applications filed after the start of the project[29]. Moreover, as initially explained, we maintained in the final dataset only H2020 projects (and related publications and patents) that started in the years 2014, 2015, 2016, so to have a sufficient time span for the generation of publications and subsequent patents (in this respect, H2020 projects started after 2016 were probably still ongoing at the time of the study). Tabel A4 presents the main records from our final dataset used in the analyses reported in Sections

[29] We decided to maintain only scientific publications for which information was available in Scopus (independently from the initial source of origin of such publication, being Cordis, the ERCEA and/or Scopus itself) so to have full bibliographic information on the publications, in particular on the publication year of the article. We also decided to maintain publications that were originally identified only in the Scopus database (and that were not available in Cordis or in the ERCEA records) only in cases where we were able to identify the PI among the authors of such publications. This is a conservative approach, given that Scopus does not always clarify the source of information used for the identification of the funding sources behind a publication.

5 and 6 of this report. This final dataset includes 6,671 ERC projects, 172,683 publications linked to them and 34,513 patent applications citing such publications, as shown in Table A4.

Table A 4 - Total number of ERC projects, related publications with Scopus Information, patents citing such publications, included in the final sample (all projects; FP7 projects; H2020 projects starting in the period 2014-2016).

Variable	Total Dataset	FP7	H2020 (2014-2016)
ERC Projects	6,671	4,556	2,115
Linked Publications	172,683	134,961	40,407
Patent applications citing linked publications	34,513	32,728	3,204

3.3 Matching between ERC PIs/Institutions and inventors/applicants of patents

Usually, a patent can have more than one inventor, but only one applicant. For each patent application identified, we extracted from Patstat all their inventors, applicants and relative information (Patstat tables *tls206, tls207, tls228, tls226*). A match was performed with the PI(s) surname, coordinators, and participants relative to the ERC-publication that generated the match. In detail, we searched for the presence of PIs both among the inventors and the applicants, while we looked for the coordinators and participants only when figuring as the patent applicant. Since personal names and names of institutions can be written in non-uniquely ways (i.e. second names, particular characters, abbreviations etc.), Patstat provides two fields with normalised names: *han_normalized* and *psn_normalized.* We used both these fields for our match.

The match of the PIs with the inventors/applicants is done by normalizing the name in the same way of the previous steps, and then searching for the exact match. This results into 3,418 different patents applications having the PIs as inventor (3,024 FP7 and 523 for H2020, meaning that some applications are in common between the two programs) and 305 patents having the PIs as applicant (FP7: 287, H2020: 21).

Identifying when the coordinator or one of the ERC participants was figuring as one of the applicants is less trivial. The name of the institution can assume different forms, appearing as an acronym rather than in its extended form, or with the name in its native language as well as translated in English. Variations are possible also due to changes in the name over the time. The match with the coordinators and participants was considered valid under two conditions. *First* at least 50% of overlap with the applicant was required, and *second*, the overlap was not to consist only of stopwords or generic words such as “university” or “institute”. For this passage we manually defined a multilingual blacklist. The match returned 3,206 patents having an ERC coordinator as applicant, from 1205 unique coordinators (FP7: 2,872, H2020: 464) and 110 patents with one of the participants of an ERC project as applicant, from 95 unique participants (FP7: 109, H2020:35). In Table A 5 we shows the match that each application has with all the four possible combinations.

Table A 5 - Number of FP7 and H2020 projects having the PI as applicant/inventor or the coordinator/participants as applicant of a patent.

FP7	**H2020**	**FP7and H2020 projects**	**FP7and H2020 Patents**	**PI as applicant**	**PI as inventor**	**Coordinator as applicant**	**Participant as applicant**
1	0	1	1	yes	yes	yes	yes
167	0	167	167	yes	yes	yes	no
0	0	0	0	yes	no	no	yes
3	2	5	2	yes	yes	no	yes
105	16	121	17	yes	yes	no	no
14	2	16	16	yes	no	no	no
7	1	8	8	no	yes	yes	yes
1,947	358	2,305	2,136	no	yes	yes	no
36	19	55	36	no	yes	no	yes
869	142	1,011	980	no	yes	no	no
4	4	8	8	no	no	yes	yes
1,062	125	1,187	1,155	no	no	yes	no
58	9	56	59	no	no	no	yes

4. ***Identifying self-reported patent applications***

In order to compare the results obtained from the method based on patents-publications-grants matches with the set of patents reported by the project's principal investigators at the ERC in the intermediate and final technical reports submitted to the Agency to illustrate the activities performed and the main outcomes achieved. The idea is to better understand how these approaches overlap or complement each other. To this purpose, we were able to use information directly provided by the ERC staff on the patents declared (or not) by each project. This internal database provided by the ERC had 2,206 intelectual property records, 1,650 declared from FP7 projects, and 556 from H2020 projects. From that list, we identified the patent application number and matched them to Patstat to gather information about application year, legal status, inventors, applicants, DOCDB family, IPC, NPL citations. From this search, we were able to identify 1,963 unique patent applications connected to H2020 and FP7 by 901 projects. From these records, 1,572 patents are from 609 FP7 projects, and 446 patent applications are linked to 292 H2020 projects.

Annex 2: Firms' patents citing publications from ERC-funded research

The table below contains the names a selected group of firms owning patents citing (in the NPL) publications from ERC-funded projects. Only the first 20 firms by number of patents citing ERC-funded research are reported in the Table.

As some companies used different names when filing their patents at the patent offices, we performed name cleaning using the "Harmonised patent aplicants' name" Table included in Patstat, 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 was beyond the scope of this study. The number of patent applications in the Table is computed using a fractional approach (i.e. in case of multiple owners, the patent is divided equally among all co-owners).

Table A 6 – Top 20 firms by number of patents applications citing ERC funded research

Firm Applicant	Number of Citing Patent Applications (fractional approach)	% (of total citing patents)	Cumul.
IBM (INTERNATIONAL BUSINESS MACHINES CORPORATION)	489.93	2.81%	2.81%
MICROSOFT TECHNOLOGY LICENSING	379.93	2.18%	4.99%
AT&T INTELLECTUAL PROPERTY I (AMERICAN TELEPHONE AND TELEGRAPH COMPANY INTELLECTUAL PROPERTY I)	227.75	1.31%	6.30%
PHILIPS ELECTRONICS	217.81	1.25%	7.55%
ASM IP HOLDING	159.83	0.92%	8.47%
GOOGLE	159.55	0.92%	9.38%
INTEL CORPORATION	146.68	0.84%	10.22%
SAMSUNG ELECTRONICS COMPANY	135.57	0.78%	11.00%
QUALCOMM	118.87	0.68%	11.68%
STATE FARM MUTUAL AUTOMOBILE INSURANCE COMPANY	105.00	0.60%	12.29%
REGENERON PHARMACEUTICALS	98.00	0.56%	12.85%
THOMSON LICENSING	95.42	0.55%	13.40%
HUAWEI TECHNOLOGIES COMPANY	93.76	0.54%	13.93%
10X GENOMICS	92.08	0.53%	14.46%
CELLECTIS	91.50	0.53%	14.99%
HRL LABORATORIES	85.58	0.49%	15.48%
LOCKHEED MARTIN CORPORATION	84.67	0.49%	15.96%
INCYTE	77.50	0.44%	16.41%
ADOBE	77.50	0.44%	16.85%
NOKIA TECHNOLOGIES	75.84	0.44%	17.29%

Annex3: Description of methods for classifying patents by technology domains

In the following sections we describe in detail the methods used to collect, process and aggregate the data used in the study to analyse patents by technology domains.

1. The IPC Classification Scheme

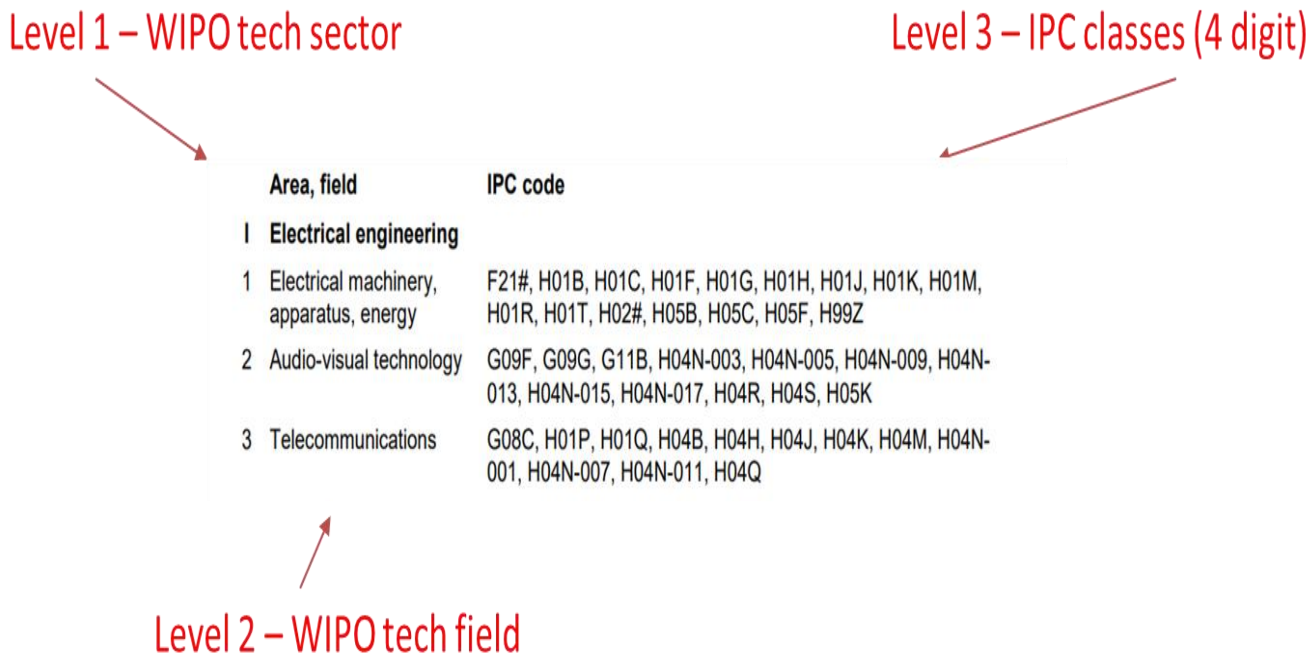
The IPC classification scheme is an internationally renowned method for classifying patents, developed with the primary aim to aid patent examiners at patent offices to perform rigorous assessment of patent applications during the examination process. It is used by patent offices from over 100 countries around the world. The purpose of the IPC system is to group patent documents according to their technical field. A patent can be included several different technological features, and therefore it can be assigned in multiple IPC classes. The IPC classification system follows a hierarchical configuration and it is at the basis of the WIPO classification scheme of technologies that we describe later. A more recent patent classification schemes deriving from IPC is the Cooperative Patent Classification (CPC), jointly developed by the EPO and USPTO and launched in 2010. The CPC is IPC-compliant, given that most of its subdivisions stem directly from current IPC entries. The CPC classification is at the basis of the concordance tables implemented by the EPO in order to identify patents related to climate-change mitigation technologies and patents related to 4IR technologies.

2. The WIPO Classification Scheme

The information provided by the IPC (or CPC) represents a first important reference for identifying patents in a speficic technological domains. However, for research and policy uses such schemes may result not easily used, as it is extremely granular and initially thougth to facilitate the work of patent applicants and examiners, rather than to provide policy insight. An alternative classification scheme that has been extensively used in both academic research and policy practice is the WIPO classification scheme originally developed by OST-INPI/FhG-ISI (Observatoire des Sciences et Technologies, Institut National de la Propriété Intellectuelle) and the Fraunhofer Institute for Systems and Innovation Research.

It is based on a list of five major technological sectors, divided into 35 technological fields categories, which are groupings of IPC subclasses (at the 4-digit level) and cover the entire IPC classification. One example of a major group is Electrical Engineering, with subset fields such as Electrical Machinery, Audio-visual technology, Telecommunications, Digital Communications, among others (see Figure A 2). This scheme is currently used, for instance, by the OECD in compling the patent counts by technology at the country level, thus facilitating the comparisons and the interpretation of results. We thus conducted the analyses of patents inspired by ERC projects in the various technology domains according to such WIPO classification scheme, referring to the 5 macro-sectors and related 35 technological. Additional information on the WIPO classification scheme is reported in the [WIPO commissioned report](#).

Figure A 2 – Example of the 3 analytical levels of the WIPO classification scheme



	Area, field	IPC code
I	**Electrical engineering**	
1	Electrical machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
2	Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K
3	Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q

3. The EPO Classification of patents related to climate-change mitigation technologies

A first focus on relevant technological trends has been conducted with respect to sustainable technologies, in particular, to Climate Change Mitigation Technologies (CCMTs) which the United Nations Framework Convention on Climate Change defines as those technologies where the human intervention has been directed to reduce the sources or enhance the sinks of greenhouse gases.

Different classification schemes have been developed in order to identify patents related to sutainable technologies. For the purposes of this study, we refer to the classification developed by the EPO and launched in 2010 (see the paper by Angelucci et al., 2018 for a detailed description of the classification scheme). Before the introduction of this scheme, patent documents relating to CCMTs were scattered throughout the IPC and the CPC and did not fall under one single classification section, as they belonged to many different technology areas. To make easier the identification of patents related to sustainable technologies, the EPO introduced a dedicated tagging scheme known as the “Y02/Y04S scheme” (and related sub-fields), which is fully integrated within the CPC.

This scheme covers seven main categories of climate change mitigation technologies, namely related to energy, greenhouse gases, buildings, industry and agriculture, transport, waste management and wastewater treatment, smart grids. Each of these categories is also subdivided into several technology fields (for a total of 28 fields). More precisely, the class Y02B includes patents relating to the integration of renewable energy sources; Y02C covers all techniques directed toward the capture, storage and sequestration of greenhouse gasses; the Y02E subclass represents the pool of renewable energy sources, technologies related to nuclear power and combustion technologies; the Y02P category covers clean industrial production processes, ranging from the primary to the secondary industry; Y02T tags patents related to electromobility, with a focus on technologies which reduce the use of fossil fuels, together with more efficient internal combustion engines, improvement of trains and aircrafts, use of light-weight and composite materials, energy recovery systems and the use of innovative maritime propulsion systems; the class Y02W is split into wastewater treatment, which is particularly focused on reduction of biogas emissions, and waste management and recycling; Y04S is the class dedicated to smart grids.

We also added a separate category, Y02A, related to technologies for *adaptation* to climate change and covers the areas of adaptation at costal zones, water supplies or resources conservation or efficiency, adaptation technologies for agriculture, forestry, livestock or agro-alimentary production, adaptation technologies aiming to protect human health and climate change resilient infrastructures.

4. The EPO Classification of patents related to the 4th Industrial Revolution

An additional focus on technological trends of high relevance for the economy and society at large has been conducted with respect to patents related to the Digital Transformation. There have been several attempts to identify patents behind digital transformation, the concept itself being not uniquely determined both in academic research and in practice. For the purposes of this study, we adopt the classification provided by the European Patent Office in the 2017 report "Patents and the Fourth Industrial Revolution. The inventions behind digital transformation". In this report, the term Fourth Industrial Revolution (4IR) is used to denote the full integration of information and communication technologies (ICT) in the context of manufacturing and application areas such as personal, home, vehicle, enterprise, and infrastructure. Adopting this perspective and the corresponding classification scheme, implemented by EPO patent examiners expert in the subject, has the advantage of relying on a pre-defined concordance table linking toghether specific 4IR technology sectors and a set of corresponding CPC patent classes.

Indeed, 4IR patented inventions have been classified by the EPO into three main sectors, each of which is subdivided into several technology fields (for a total of 18 fields). Concerning the 3 sectors: a) Core technologies (Hardware, Software and Connectivity) include technologies that allow to transform any object into a smart device connected via the internet. B) Enabling technologies (Analytics, Security, Artificial intelligence, Position determination, Power supply, 3D systems, User interfaces) are used in combination with connected objects; C) Application domains (Home, Personal, Enterprise, Manufacturing, Infrastructure, Vehicles) identify contexts where the potential of connected objects can be exploited. Further details on such methodology and the related concordance table containing around 320 CPC field ranges in all technical areas with their respective 4IR technology fields can be obtained in the original report from the EPO.

Annex 4: Regression analyses on the relationship between patent citations and scientific citations received by a project's publications

In the following Table we performed a set of regression analyses based on the final sample of ERC funded FP7 and H2020 projects used in the report in order to analyze the probability of an ERC project to generate publications cited by subsequent patent applications (Models 1 and 2). Models 3 and 4 consider, as dependent variable, the number of patent applications citing (in the NPL) the project's scientific publications.

The explanatory variable of primary interest for this analysis is represented by the number of citations received in Scopus by the publications of the ERC project (in order to control for time effects and to normalise for the comparison of projects across different time periods, in Model 1 and 2 we consider only scientific citations received for each publication in a 3 year time window from the publication year, whereas in Model 3 and 4 we consider a 5 year citation window). We then include in the regression models dummy variables to control for program type (FP7 vs H2020), ERC project type, scientific domain of the project, starting year of the project.

The results of the regression models confirm the positive and statistically significant relationship (at the 1% level) existing between the scientific excellence of a research projects (as captured by citations received from scientific literature) and its technological influence (as captured by citations obtained from subsequent patents).

Table A 7 - Regression analyses on the relationship between patent citations and scientific citations received by a project's publications

	Project's likelihood to be cited by patents		**Number of patents citing the project**	
	Probit		**Nbreg**	
	M1	M2	M3	M4
Scopus Citation 3 years	0.0004*** (0.0001)		0.002*** (0.0001)	
Scopus Citation 5 years		0.0003*** (0.0001)		0.001*** (0.0001)
Program	Yes	Yes	Yes	Yes
ERC Type	Yes	Yes	Yes	Yes
Project Sector	Yes	Yes	Yes	Yes
Start Year	Yes	Yes	Yes	Yes
N Observations	*6,671*	*6,671*	*6,671*	*6,671*
R^2	*0.2162*	*0.2221*	*0.0895*	*0.0924*