European Commission

JRC TECHNICAL REPORTS

Final report on the collection of patents and business indicators by economic sector: Societal Grand Challenges and Key Enabling Technologies

Collection and analysis of private R&D investment and patent data in different sectors, thematic areas and societal challenges

Frietsch, R. / Kladroba, A. / Markianidou, P. / Neuhäusler, P. / Peter, V. / Ravet, J. / Rothengatter, O. / Schneider, J.

Joint Research Centre

Contact information
Dr. Rainer Frietsch
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48
76139 Karlsruhe, Germany
Email: rainer.frietsch@isi.fraunhofer.de
Tel.: +49 721 6809-197

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

JRC/BRU/2014/J.6/0015/OC

JRC109299

ISBN 978-92-97-76847-7 doi:10.2760/39818

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Frietsch, R. / Kladroba,A. / Markianidou, P. / Neuhäusler, P. / Peter, V. / Ravet, J. / Rothengatter, O. / Schneider, J., *Final report on the collection of patents and business indicators by economic sector: Societal Grand Challenges and Key Enabling Technologies*, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-97-76847-7, doi:10.2760/39818, JRC109299.

Contents

Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

The estimates for employment provided in this report are based on relatively strong assumptions. Whereas the strong correlation between patents and R&D expenditures justifies the approach used in this report for estimating BERD, the connection between patents and employment is less evident. Readers should interpret and use these estimates with caution.

Authors

Frietsch, Rainer

Kladroba, Andreas

Markianidou, Paresa

Neuhäusler, Peter

Peter, Viola

Ravet, Julien

Rothengatter, Oliver

Schneider, Julia

Executive Summary

- The aim of the "Innovation Union" strategy is to create jobs and welfare based on innovation. Horizon 2020 as the financial instrument implementing the Innovation Union puts forward the idea of a mission orientation.
- This mission orientation is first of all based on Societal Grand Challenges (SGCs) and Key Enabling Technologies (KETs), which are crucial technologies that built input or precondition for the competitiveness in these Societal Grand Challenges.
- In this report, information is provided on business R&D expenditures and patenting structures within Key Enabling Technologies and Societal Grand Challenges.
- The main challenge hereby was to estimate BERD, value added and employment at the technological level of KETs and SGCs as these indicators are only available at the sectoral (NACE 2-digit) level.
- The employed method uses the (weighted) distribution of technology-specific patents per sector to re-allocate BERD, employment and value added by technology fields. While the strong correlation between patents and R&D expenditures is a proof of the use of this approach for BERD, the connection between patents and employment as well as value added is much smaller. For these latter two indicators the estimated values are therefore to be cautiously interpreted. Yet, technology specific data on value added and employment is not available. Variations across countries, could partly be taken into account by employing four country groups.
- The descriptive analysis examines business R&D expenditure, employment, value added and (transnational) patents of each KET and SGC in terms of level, trend, intensity and specialisation. The multivariate analyses deal with relationships between R&D expenditures, patents and value added.

KETs

- Largest figures for R&D expenditure in the EU28 are found in Advanced Manufacturing Technologies (AMT), followed by Nanotechnologies and Advanced Materials. These three KETs are also the largest in terms of employment and value added.
- Largest KETs in terms of patents in the EU28 are AMT and Advanced materials, followed by Micro- and nano-electronic technologies and Photonics. These latter two KETs also present the largest number of patents per unit of R&D spent.
- Specialisation indices (based on R&D and patents) show a high specialisation of EU28 in AMT and low specialisation in biotechnology.

SGCs

- The transport SGC stands ahead of all other SGCs for all indicators in level. Health is ranked second for R&D, employment and value added.
- Along with transport and health, food also presents high figures in terms of patents compared to other SGCs.
- While climate is the smallest SGC based on the different indicators examined, it also presents the largest number of patents per euro of R&D spent.
- The EU28 does not present strong specialisation in a particular SGC compared to other countries, meaning that the above results represent a balanced picture in comparison with other countries.

Multivariate Analyses

- In order to assess the relationship between Business Enterprise Expenditure on R&D (BERD), patent filings and value added within KETs and SGCs, a series of panel regression models (sequential modelling approach) was calculated.
- BERD is positively related to patents and patents have a positive effect on the economic output measured by value added in the fields of KETs.
- In the case of SGCs, the relationship is more ambiguous and direct connections cannot be found in the same way as BERD does not explain the number of patent filings in a statistically significant manner.
- An overall global superior position in all KETs fields and a worldwide market leadership in these technologies seem neither achievable nor desirable for Europe. Based on existing literature, a selective strategy in KETs, i.e. specialising only on the fields where Europe is already strong, and a general strategy in SGCs seems much more favourable and productive.

It is meaningful and reasonable to follow a mission-oriented innovation policy approach as it is done under Horizon 2020 that allows both funding individual technologies and technological solutions.

1 Introduction

Within the Europe 2020 Strategy, the 'Innovation Union' (IU) flagship initiative is a cornerstone policy strategy. In order to create growth and jobs Europe should take the lead in certain technological areas and innovation-driven sectors: "*By improving conditions and access to finance for research and innovation in Europe, we can ensure that innovative ideas can be turned into products and services that create growth and jobs*." (European Commission 2013, page 8).

The Innovation Union's main tool is the EU Framework Programme for research and innovation. The current Programme – Horizon 2020 – differs slightly from earlier research programmes given its stronger turn on mission orientation. Under the umbrella concept of "societal grand challenges" (SGC) solutions and/or applications are in the focus instead of singular and individual technologies. The SGCs guide decision making and agenda setting in terms of science funding and innovation policy. The European Union formulated seven such challenges including energy, mobility, climate, health, food, and security.[1] The challenges are (effectively but maybe not intentionally) defined according to existing strengths in Europe and the intention is to keep or even enhance them.

While the SGCs serve as guiding main objectives, a number of technological fields are specified in the various work programmes. They are key to address the challenges and are seen as important input or preconditions for the European competitiveness in these Societal Grand Challenges. While there are the societal challenges, the European Union's H2020 programme is also focussing on what is called 'key enabling technologies' (KETs) as part of 'Leadership in Enabling and Industrial Technologies' (LEIT).[2] Among these are advanced computing, robotics or new electronic components. As a particular subarea of the LEITs, the Key Enabling Technologies (KETs) have already played an important role in FP7. KETs comprise of nanotechnology, photonics, industrial biotechnology, advanced materials, micro- and nano-electronics.[3] Advanced Manufacturing Technologies (AMTs) are seen as a cross-cutting area which support the other five KETs and therefore play a key role within KETs. AMTs cover process technologies that are used to produce any of the other five KETs. This includes production apparatus, equipment and procedures for the manufacture of specific materials and components (IDEA Consult et al. 2015).

Key Enabling Technologies are expected to be of high relevance in a number of future products and processes.[4] And they are equally important in context of achieving the Societal Grand Challenges. For those, they serve as a direct input or provide basic competences and technological solutions.

In this context, a number of studies addressed KETs, but some also individual SGCs. They were mainly inspired by a research and innovation related "European paradox". It states that Europe has large competences in research and technology but performs below average in terms of commercialisation and exploitation. Between these two sides is the "valley of death", where problems in converting knowledge into marketable products and services prevails. This phenomenon seems to be explicit for KETs (Larsen et al. 2011). It is a problem identified by several European countries and addressed by national policies. KETs as well as the broader group of LEITs were identified as important for the economic and innovation-based development (DTI et al. 2013; Reiß et al. 2016).

The research and innovation policies for KETs and SGCs focus on the positioning and competitiveness of Europe within these fields. With respect to KETs in Europe, Neuhäusler et al. (2015) conclude an excellent position in the subfield of Advanced Manufacturing Technologies (AMTs), while in the other subfields, the European position is

[1] The seventh challenge is "Inclusive, innovative and reflective societies". The challenge is not analysed in this report since its technological foundations are rather limited. It thus cannot be taken into account with the tools and approach applied here.

[2] LEIT includes KETS (key enabling technologies), space and ICT.

[3] Under FP7 industrial biotechnology was not included under KETs.

[4] https://ec.europa.eu/growth/industry/key-enabling-technologies_en

rather weak. In terms of the current global position of the EU-28 countries in KETs and – forward-looking for SGCs in the year 2020, Reiß et al. (2016) provide evidence for a solid and good positioning in most areas. Since also several non-European countries focus on SGCs, Reiß et al. project for 2020, that the position might be challenged - especially from companies located in Asia. If Europe continues with "business as usual" they conclude that the relative position of the EU in an expected growing market could nevertheless weaken. As the overall market grows, the absolute levels of Europe might also increase – thereby creating jobs and growth. The overall market may grow faster than Europe's relative position weakens. In other words, it is expected that the cake becomes bigger, but the share of the cake for Europe could be smaller. Still, the smaller share might be larger than the one from a smaller cake – hence, the net effects in growth and jobs. The positioning of Europe in SGCs might even get better as these are already outstanding areas of European strengths, while KETs start from a weaker position. European policy makers may be aware that global leadership in KETs is unlikely to be achieved, but specialisations in subfields which are of particular relevance to the performance within the SGCs seem to be advisable. "Strengthening the strengths" within those parts of KETs directly relevant as enablers for the Societal Grand Challenges is thus among European policy attention.

This report tries to provide empirical evidence for the European activities in KETs and SGCs. It analyses the relationship between R&D expenditures and structural changes in KETs and SGCs. After describing the data sources and the methodological details in chapter two, descriptive statistics on Business Expenditures on Research and Development (BERD), on transnational[5] patent applications (TNL), value added (VA) and employment (EMP) are provided in chapter three. Based on an integrated panel data (consisting of the aforementioned indicators), multivariate analyses of the relationship between those indicators are carried out in chapter four.

[5] Transnational patent applications are patent families with at least one member at the EPO or via the PCT procedure at the WIPO. For a definition and a methodological discussion of transnational patents see Frietsch and Schmoch (2010). With transnational patents the home advantage of certain countries at certain patent offices can be overcome and their technological output can be made comparable. At the same time, this perspective also allows to have consistent and continuous time series of the data up to the latest available year of priority filings (18 months before the current date of analysis). For this report, using the PATSTAT version of September 2015, we have patent data available and comparable up to the year 2013.

2 Methods

In this chapter, we will first of all describe the data sources used for our analyses. In addition, the strategy for imputing missing data in the BERD, value added and employment statistics will be explained. In a third step, the conversion of the data that is only available at the sectoral level, i.e. BERD, value added etc., to fields of technology (KETS, SGCs) will be discussed. This is a crucial step within this project as it allows us to combine the data on BERD, value added and employment with the patent data to perform further analyses. Besides providing descriptive statistics on these indicators, we have additionally calculated multivariate models on the relationship between BERD and patent filings as well as patent filings and value added within KETs and SGCs. These will also be described below.

2.1 Data sources and definitions

For this project, several data sources have been used. These are described in more detail below. Due to the large coverage in terms of countries, years and sectors, however, gaps within the data still remain. In the following chapter, we will therefore additionally describe our strategy to impute these values in order to gain a comprehensive set of indicators with the largest possible coverage.

BERD, value added and employment

Economic data on business R&D expenditure, value added and employment were collected by NACE Rev.2 from Eurostat, OECD and national authorities. The definitions for these indicators are the following:

- BERD: Business enterprise expenditure on R&D (BERD) covers R&D activities carried out in the business sector by firms and institutions active in R&D, regardless of the origin of the funding. The business enterprise sector includes: i) all firms, organisations and institutions whose primary activity is the production of goods and services for sale to the general public at an economically significant price, and ii) the private and not-for-profit institutions mainly serving them (source: OECD).
- (Gross) value added: Gross value added is the value of output less the value of intermediate consumption; it is a measure of the contribution to GDP made by an individual producer, industry or sector; gross value added is the source from which the primary incomes of the SNA are generated and is therefore carried forward into the primary distribution of income account (source: OECD and Eurostat).
- Employment: It covers all persons engaged in some productive activity (within the production boundary of the national accounts). Employed persons are either employees (working by agreement for another resident unit and receiving remuneration) or self-employed (owners of unincorporated enterprises) (source: Eurostat).

Patents

The patent data for the study were extracted from the "EPO Worldwide Patent Statistical Database" (PATSTAT, April, 2016 version). PATSTAT contains information on published patents from more than 80 patent authorities worldwide, dating back to the late 19th century. It includes all the information stated on a patent application, i.e. application authorities (patent offices), several patent relevant dates (priority, filing, and publication date), the type of application (invention patent, utility model, etc.), inventor and applicant addresses, patent families, patent classifications (IPC and ECLA), title and abstract of a patent filing, technical relations and continuations, as well as citations of patents and of non-patent literature. We used the automated tool developed by the K.U. Leuven for name cleaning and applicant name harmonisation (Du Plessis et al. 2009; Magerman et al. 2009; Peeters et al. 2009). The patents in our analyses are counted according to their year of worldwide first filing, which is commonly known as the priority year. This is the earliest registered date in the patent process and is therefore closest to the date of invention. To allow international comparability of technological/inventive capabilities, we follow a concept suggested by Frietsch and Schmoch (2010) called transnational patents, which is able to overcome the home advantage of domestic applicants, so that a comparison of technological strengths and weaknesses becomes possible – beyond home advantages and unequal market orientations. In detail, we count all patent families that have at least one PCT or EPO application as a member.

Patent definitions

The Societal Grand Challenges were defined in terms of technologies. For the definition of the content and the demarcation of the Societal Grand Challenges as such, we have relied on the legal base of the First Working Programme of Horizon 2020. In a second instance, a number of interviews were conducted with representatives of the Commission Services in order to generate the definition of SGCs based on the International Patent Classification (IPC). A detailed description of the methodologies and the underlying definitions can be found in Frietsch et al. (2016).

Key Enabling Technologies are core areas for achieving the strategic goals addressed in EU 2020. In the case of KETs, we resort to the most recent definition that has been developed by the KETs Observatory (IDEA Consult et al. 2012). The definition is based on the International Patent Classification (IPC).

2.2 Imputation Strategy

Missing values for BERD, value added and employment and the level of 2-digit NACE Rev.2 sectors have been addressed by several imputation steps and by converting data from other existing sectoral classifications (in particular NACE Rev.1.1, ISIC Rev.4 and ISIC Rev.3). This strategy takes into account that combining sectors, years and indicators entails vast numbers of individual values, which we cannot reasonably expect national offices to provide data for. The following imputation methods have been applied to the data in a stepwise process from method 1 to method 8:

- Method 1 - calculating the difference: Should all values except one be available for an aggregate, the missing value is to be determined by calculating the difference between the values in t-1 and t+1.
- Methods 2 to 4 - interpolation: The missing value is determined by means of interpolation between an earlier and a later value.
- Method 5 - shares: The missing value is determined as average proportion of given values from other years of the next higher aggregation level.
- Method 6 - calculating the difference based on imputed values: Should, after imputation, all values except one be available for an aggregate, the missing value is to be determined by calculating the difference.
- Method 7 - Swiss imputation method: The missing value is determined by means of interpolation between an earlier and a later value taking into consideration the gross domestic product (GDP), and
- Calibration: The calibration factor used accrues from the ratio between the sum of the missing data and the sum of the imputed data.
- Method 8: for value added and employment, in case the sector assignments in NACE Rev. 2 and NACE Rev. 1.1 were unambiguous (according to the correspondence table provided by Eurostat[6]), NACE Rev. 1.1 data were imputed if available in the respective country/year combination (the same process was performed for ISIC Rev.3

[6] http://ec.europa.eu/eurostat/web/nace-rev2/correspondence_tables

and ISIC Rev. 4, which are identical to NACE Rev. 1.1 and Rev. 2 at the two digit level). This imputation method only concerns about 0.7% of data points presented in this report.

A more detailed description of the imputation methods and steps can be found in the methodological report that deals specifically with these questions.

2.3 Conversion of sectoral data to technology fields (KETs, SGCs, WIPO35)

One of the major challenges within this project was to estimate BERD, value added and employment by KETs, SGCs and the WIPO35 list of technology fields, as these indicators are only available at the sectoral (NACE 2-digit) level. To do this, we resorted to a matrix of (transnational) patent filings by NACE sectors and KETs/SGCs/WIPO35, which was generated by linking the 2013 EU Industrial R&D Investment Scoreboard (Hernández et al. 2014) with PATSTAT at the level of companies/patent applicants. For the match of PATSTAT and the 2013 EU Industrial R&D Investment Scoreboard, a string matching algorithm based on the Levenshtein distance was applied after a cleaning of company/applicant names (for more details on the general method compare Neuhäusler et al. 2016; Neuhäusler et al. 2015). Based on this matched data set, we were able to generate probabilistic concordance schemes based on the shares of patents by sectors in each of the fields of technology.

Since we cannot assume that these structures are equal across countries, we calculated four matrices, one for North America, Asia, North-Eastern Europe and South-Eastern Europe each. For countries from the rest of the world that did not belong to one of these groups, i.e. Brazil, the matrix for South-Eastern Europe was applied.

These matrices of patent shares were then applied to relocate the BERD, value added and employment data by KETs, SGCs and WIPO35.[7] If, for example, the patent shares show that patent filings from one sector are split up between 50% to KETS1, 30% to KETS2 and 20% to KETS3, the BERD, value added and employment data were split up accordingly. The relocation covers all sectors, i.e. manufacturing and services, implying that also BERD, value added and employment data from service sectors were relocated to the level of technology fields. Since the KETs and SGCs do not cover all R&D expenditures/patents in a given country, a respective share for "non-KETS" was applied. If there were no patents within a sector at the 2-digit level, the values from the 1-digit level were used.

Before the data could be relocated, however, it had to be weighted, since it cannot be assumed that a patent has the same "price" in terms of R&D input in each of the fields, e.g. a patent in the fields of pharmaceuticals can be assumed to consume larger amounts of R&D expenditures than a patent in the field of other consumer goods. To generate such a weighting scheme, we had to rely on an external source to not run into the problem of a circular logic. We thus resorted to an estimation of R&D intensity, defined as the amount of R&D expenditure per patent, by Schmoch and Gauch (2004).[8] In case more (or less) R&D expenditures per patent are necessary in a given technology field, this technology field is assigned a higher (lower) weight within the concordance matrix.

[7] These calculations are based on the assumption of a high correlation between R&D expenditures and patents. We have performed the same calculations for value added and employment, though is has to be taken into account for the interpretation that the correlation between patent filings and these two indicators is not as high as in the case of R&D. In order to balance for the different industry structure of countries, we have created four matrices for four country groups, i.e. South-Eastern Europe, North-Western Europe, North America and Asia. Yet, it is assumed that these structures (and correlations) remains similar over time.

[8] The list of (Schmoch, Gauch 2004) includes only 19 technology fields, i.e. the 19 fields were assigned several times to the fields in the list of the 35 technology fields (Schmoch 2008). In case of multiple assignments, each of the fields of the list of 35 was assigned an equal weight according to the list of 19 fields. To account for KETs and non-KETS as well as SGCs and non-SGCs, we applied weights from similar fields from the WIPO list. For non-KETs and non-SGCs, the average weight of the remaining WIPO35 fields was applied.

The weighting scheme was then applied to our concordance matrix before relocating BERD, value added and employment by technology fields. This leads to a final panel data set where all indicators, i.e. patents, BERD, value added and employment are available for the KETs, SGCs and WIPO 35 by country and year. This data set was used for all further analyses.

2.4 An attempt to validate the conversion of BERD into technologies

It is almost impossible to validate the results described above of converting BERD (and VA and employment) by sectors into BERD by technologies described above. This is because such data is rarely available; even within companies/firms, the data is hardly known as conventionally either total R&D expenditure or R&D by projects or departments is documented. An aggregation/summation at the level of technologies is not usually made as this is not a level of analysis/management in firms and is (usually) not requested in regular R&D surveys. Besides, if the data by technologies were available, we would not need to suggest a method to estimate it.

Insights from a previous study

There are only a few instances of available R&D data by technologies/disciplines. In a previous project on behalf of the German Federal Ministry of Education and Research (BMBF), Stifterverband and Fraunhofer ISI developed the conversion method based on the German R&D survey (Frietsch et al. 2014; Neuhäusler et al. 2016). In that project, we used data from the pharmaceutical industry association and for nanotechnology from a survey by the BMBF to calibrate the conversion matrix. The sources state that, in 2009, industry spent about 4 billion Euros on pharmaceutical research in Germany and almost 1 billion Euros on nanotechnology. The other technology fields - according to the technology definition by WIPO (Schmoch 2008) - were calibrated respectively. This previous exercise resulted in estimated R&D expenditures on transport technologies (WIPO field No.32) of about 13 billion Euros in Germany. Here, we overestimate the R&D expenditures for transport technologies to about 19 billion Euros in Germany in 2009. In consequence, we underestimate the expenditures for pharmaceuticals and also for biotechnology, while for nanotechnology the estimation is almost the same as in the previous study.

R&D expenditures by field of science published by Eurostat

At a workshop in our project, an alternative data source for the calibration of the data was suggested: The Eurostat database on R&D by field of science, which was collected for a small number of countries and for limited years. This data source relies on the national R&D surveys and differentiates the R&D expenditures by scientific disciplines. However, this is not exactly a differentiation by technologies as it is intended here, which is one reason for discrepancies. Data is available for some countries and years for (industrial) biotechnology, nanotechnology or also for disciplines called mechanical engineering or electrical engineering - titles that we can also find in our lists of fields. However, the definitions of these disciplines/fields are not known so that we cannot assess whether any deviations between our data and the Eurostat data are due to the estimation or to a different definition/demarcation of the field. In addition, data is not available for the larger European countries like the UK, Germany or France, but only for Croatia, Latvia, Hungary and Portugal. Some data points are also available for Romania and Slovakia. For the countries for which data is available, our estimations for nanotechnology only deviate by about 1-2 million Euros (in relation to total volumes between 1-4 million per year) in most of the available years, but the data provided by Eurostat is rather volatile, while our estimates are more stable. In the case of biotechnology, our estimates are lower than Eurostat data for Hungary, but higher for Portugal. For Latvia, our results are similar with about 200,000 Euros compared to 140,000 Euros reported by Eurostat. In the case of Slovakia, we deviate strongly from the figure set by Eurostat. We estimate 400,000 Euros R&D expenditure in industry for biotechnology in Slovakia, while Eurostat reports about 4 million Euros. In sum, the deviations do not follow a clear structure and can therefore not be used, either for calibration or for a validation/invalidation of our estimates.

Alternative patent data

As described above, we employed regional conversion matrices for northern and western Europe, for eastern and southern Europe, for North America and for Asia, based on the average patent profiles of the countries assigned to these regions. However, country-specific idiosyncrasies might lead to considerable deviations from the "real value" in a few countries. In addition, we used transnational patent applications here unlike in the previous study, where we used national patent applications to better reflect the total technological output and not only the output that is intended to be relevant for international markets. The reason is a practical one in this case: national patent applications were not available in this dataset and they are also not easy to collect. For example, the fact that a number of smaller countries - among them also technology-oriented countries like Switzerland, the Netherlands or Belgium - do not have a large national market for their technological inventions and therefore file considerable numbers of patents only abroad (e.g. in Germany or the USA) or at the EPO makes this even more complicated. In addition, there are not only big differences between countries, but also within countries in the case of technology fields. For example, in Germany, about 90% of the national applications in pharmaceuticals are also filed internationally (transnational patent filings), while this figure is closer to 50% in mechanical engineering or transport. At the same time, all this might also be part of the explanation of why, for some countries, the transnational patent count covers larger shares of their patent output, while for others - like Germany - this is less so.

Alternative estimation based on German information

For reasons of comparison and as a kind of benchmark, we estimated the R&D expenditures for all countries taking the above mentioned issues into account. As we have the relevant information only for Germany, we focus on German data/structures, but apply these to all countries. First, we estimated the share of transnational over national patent applications for each technology field - 35 WIPO fields, 6 KETs and 6 SGCs - and "weighted" the transnational patents accordingly. The aim was to overcome the fact that, in some technology fields, smaller shares of total technological output are covered by transnational filings. As we link total national R&D expenditures by companies to patent data, it is justified to use the total technological output represented by patents.

Second, we used the R&D per patent weighting scheme of Germany using total national filings and not only transnational patents. Thus, we essentially employed a different weighting scheme, where - for example - transport is less cost-intensive compared to the transnational perspective. In effect, transport patents are assigned a lighter weight and therefore we estimate lower R&D expenditures for transport.

Figure 1 to **Figure 4** depict the comparison of the estimation method using transnational patent applications and the international weighting scheme (international method) versus the "adapted" patent shares taking into account the different relations between transnational and national patent applications in Germany and employing the national weighting scheme (national method). The absolute R&D expenditures for the EU-28 countries are compared for KETs and SGCs. The left panel shows the results of the international method, while the right panel shows the results of the national method. When the latter is applied, a devaluation of Advanced Manufacturing Technologies results - a field that originates in mechanical engineering, where the shares of international over national filings are rather low in Germany - and higher values for Advanced Materials - again a field where high shares of total patent applications are also filed internationally. We estimate about 15 billion Euros R&D expenditure by industry in 2012 in the EU-28 countries for Advanced Materials based on the national method, while we only estimate about 6 billion Euros based on the transnational method. In Advanced Manufacturing

Technologies, estimates are 13 billion Euros using the transnational method compared to 10 million using the national method.

When looking at the estimated R&D expenditures in the SGCs, it becomes obvious that the differences for transport are considerable. With the transnational method It is 85 billion Euros are estimated in 2012 compared to 63 billion Euros using the national method. The national method results in higher values for the other SGCs, especially health.

Outlook to future validations

For future analyses, a validation and calibration procedure might become available. The Stifterverband explicitly asked the firms in the German R&D survey in 2015 to report their R&D expenditures by technology field as well. This data will soon be available. It might be possible to directly compare the method developed in the previous project. A second validation possibility is to use data from Statistics Canada, which receives procedural data from the R&D tax exemption forms from Canadian companies. In these forms, companies indicate the technologies in which they invested R&D expenditure. A reproduction of our method for Canada, developing a Canadian matrix and/or applying the North-American matrix developed here, might allow a comparison with the micro data (or the national aggregates) of Canadian companies.

Especially the exercise of the alternative estimation based on the German weighting scheme and the German relation between transnational and national filings suggests two things that might improve the estimates, but that also considerably increases the necessary efforts. First, country-specific matrices instead of regional matrices might be more appropriate for a number of countries to take idiosyncrasies like different national markets into account. This would also better reflect differences between technology fields. Second, applying national weighting schemes of R&D per patent instead of one scheme for all might also improve the estimates. However, this weighting scheme must be independent of the estimates themselves so that no circular argumentation occurs. We suggested the concordance matrix developed by Schmoch and Gauch (2004), which could in principle be applied to any country, as long as the R&D expenditures and the number of patents are available.

Figure 1: BERD (in EUR million) for KETs - average international weights

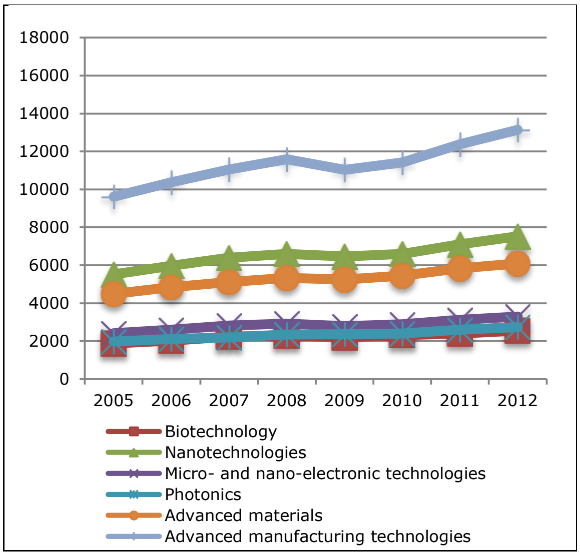


Figure 2: BERD (in EUR million) for KETs - national weight (Germany)

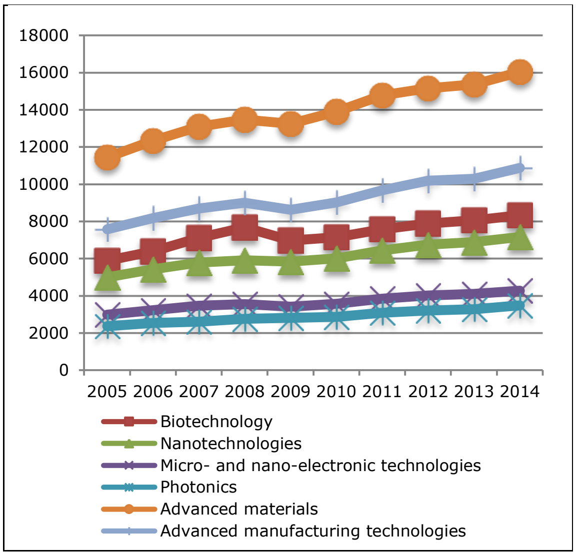


Figure 3: BERD (in EUR million) for SGCs - average international weights

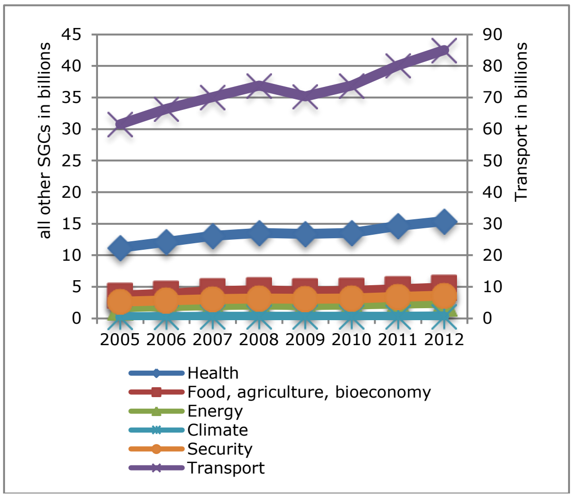
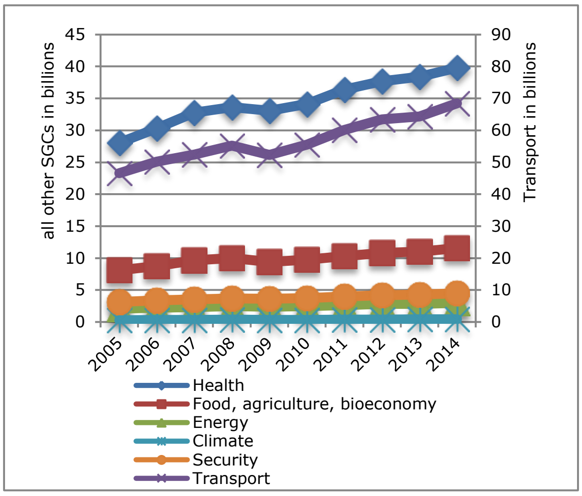


Figure 4: BERD (in EUR million) for SGCs - national weight (Germany)



Sources: OECD - MSTI; EPO - PATSTAT; Fraunhofer ISI calculations.

In sum, there is no right or wrong approach in this analysis that can be easily identified. The alternative estimations and the comparison with external data sources did not provide a clear and unique answer. Sectoral approaches also are not directly comparable as they take all R&D expenditures by a certain company into account for one given sector. For Daimler, for example, a sectoral approach would assign all R&D expenditures to the transport sector, while we only focus on Daimler's R&D expenditures for transport technologies and exclude, for example, textiles, finishes etc.

In consequence, neither validation nor invalidation was possible based on the existing information. We decided to use the transnational method and report and discuss its results throughout the rest of this report. However, there seems to be room for improvement to the estimations when many countries are involved and compared. The analyses for Germany showed that a more educated estimate is possible by applying differentiated data.

2.5 The influence of patent office definitions on country rankings

In this chapter, we discuss the differences in patent filing figures depending on the chosen definition of the patent office. In the dataset compiled within the project, patent figures are available for five different patent office definitions: the European Patent Office (EPO), the U.S. Patent and Trademark Office (USPTO), the World Intellectual Property Organization (WIPO) (PCT) filings, Transnational, and IP5 as defined by OECD (Dernis et al. 2015).

The reason for the delivery of patent data for all these different offices is a different analytical perspective depending on the selected office as well as comparability of countries, overcoming the so called home advantage. The majority of patent applicants file their patents at their national office to achieve patent protection on their domestic market. After the priority filing at the national office, they still have 12 months to decide whether to file their patent internationally as well, according to the Paris Convention.[9] This is especially true for large countries, where the domestic market promises a large sales potential in a homogeneous economic area, e.g. the U.S. or China. Yet this home advantage induces a bias in statistical comparisons of the technological competitiveness of countries using patent statistics. For example, comparing the USA and France at the USPTO would result in an unbalanced analysis concerning their technological competitiveness. On the other hand, the analysis of USPTO data might be of interest if the research question focuses on the US-American market.

The concept of transnational patent filings was suggested by Frietsch and Schmoch (2010) to counteract the problem of unbalanced comparisons when assessing the technological competitiveness of nations. This concept is able to overcome the home advantage of domestic applicants, so that it becomes possible to compare technological strengths and weaknesses beyond home advantages and unequal market orientations. In more detail, all PCT applications are counted, whether transferred to the EPO or not, and all EPO applications without precursor PCT application. Double counting of transferred Euro-PCT applications is thereby excluded. In simple terms, all patent families with at least one PCT application or EPO application among the family members are taken into account.

The IP5 Offices, on the other hand, is a forum established in 2007. It consists of the five largest IP offices in the world; namely the Japanese Patent Office (JPO), the EPO, the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO) and the USPTO. The IP5 aims to improve the quality and efficiency of patent search and examination activities and to promote work sharing among the five offices (Japan Patent Office (JPO) 2017).

Since the IP5 Offices account for approximately 80% of all patent applications filed worldwide and they offer an opportunity for international comparisons within patent statistics, the OECD and the European Commission have proposed three different search strategies to count IP5 patent filings (Dernis et al. 2015), also with the intention of overcoming the home advantage and finding a substitute for the former Triadic patent count (Grupp et al. 1996):

— **Definition 1:** Families of patent applications with members filed at one or more IP5 offices, including single filings. This implies that applications filed only at one of the IP5 offices, i.e. EPO, USPTO, JPO, KIPO and SIPO, are taken into account.

— **Definition 2:** Families of patent applications with members filed at at least one of the IP5, excluding single filings. This implies that applications filed at only one of the IP5 offices, i.e. EPO, JPO, KIPO, SIPO and USPTO, are considered only if another family member has filed at any other office worldwide (anywhere in the world, not necessarily at another IP5 office).

[9] This also works the other way around. In case of an international first filing, patent applicants also have 12 months to decide whether to file their patent directly at their national office or have their patent forwarded to it after the regional phase.

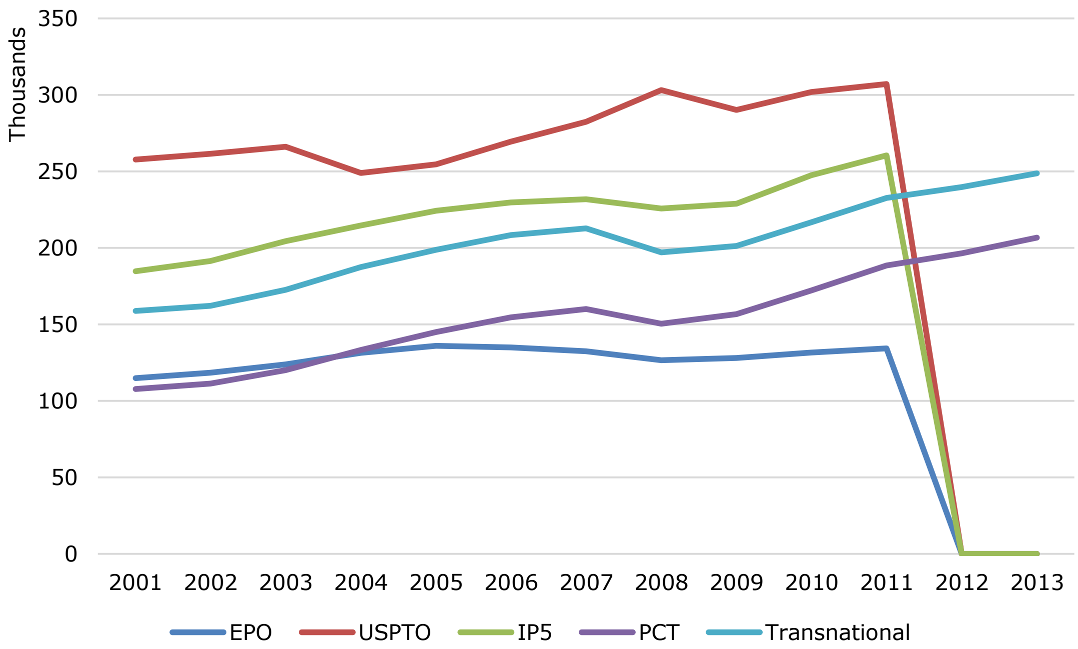
— **Definition 3:** This is the most restrictive definition, because families of patent applications are considered only if family members have been filed at at least two IP5 offices. For instance, patents filed at USPTO will be considered only if an equivalent filing has been made at any of the remaining IP5 offices. This is irrespective of whether equivalent applications also exist at non-IP5 offices.

Within the course of this project, we compiled data based on IP5 definition 2, which represents a compromise between the most restrictive definition 3 and definition 1, which includes single filings.

Comparison of the definitions

A comparison of the absolute filing figures by patent office is depicted in **Figure 14**. These worldwide figures deliver two basic messages. First, there are large differences with regard to the absolute filing figures per patent office, although the general trend – i.e. growth in the number of filings over time – is similar across the different definitions. In terms of absolute filing figures, the USPTO is the largest patent office with more than 300,000 filings in 2012.

Figure 5: Worldwide number of patent filings by patent office



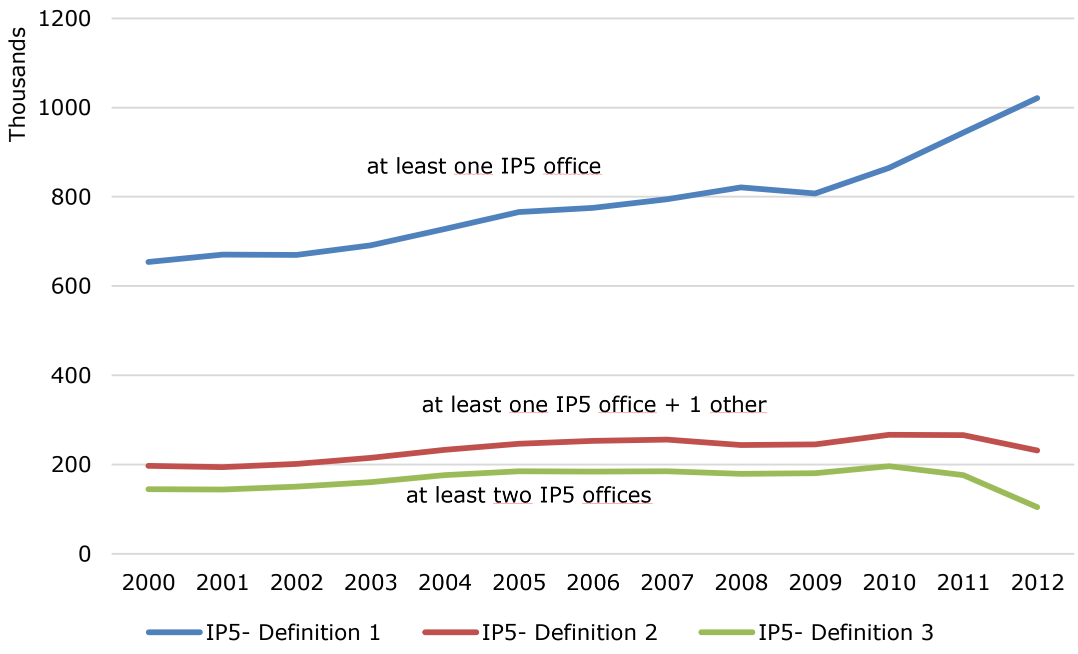
Source: EPO – PATSTAT; Fraunhofer ISI calculations.

Note: Incomplete figures for certain years are not included in the dataset, which is why some data for more recent years drop to 0 in **Figure 14**.

The IP5 definition 2 has the second largest number of filings, followed by the transnational approach, PCT filings and the EPO. A second effect visible in the graph is the periodic coverage of the indicators. All definitions that include PCT filings, i.e. PCT and transnational, provide more up-to-date figures. Usually, patent filings are published 18 months after being filed at the respective patent office. This is true for the EPO, the USPTO and most other offices worldwide. The PCT procedure, however, gives patent applicants the opportunity to request a preliminary search report, which results in an extension of the entry into the regional/national phase 12 months after priority filing. In sum, it takes a PCT filing 30 months to enter the process at a national office like the USPTO or a regional office like the EPO. This induces a statistical delay in the entry of PCT filings, leading to incomplete figures in more recent years. For example, about 60%

of the patent filings entering the EPO – this figures may be even higher depending on the country of applicant/inventors– come via the PCT route (Frietsch et al. 2013). When including PCT filings in the definition of patent analyses, however, this statistical delay does not arise as the filings are counted directly at the WIPO; there is no effect of the 12 months delayed forwarding.

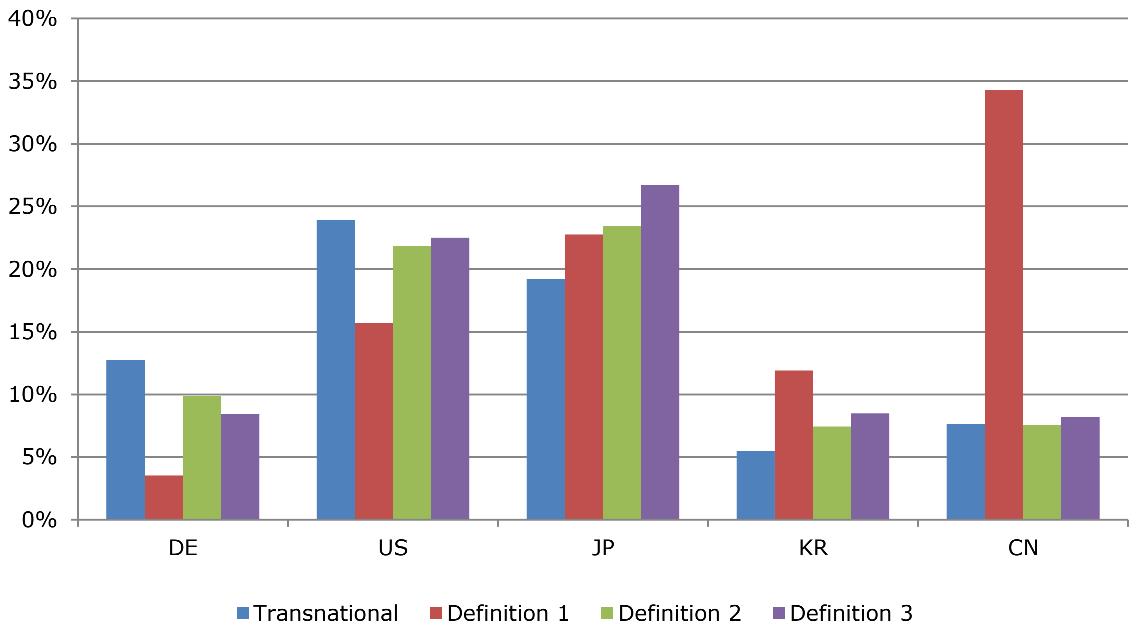
Figure 6: Differences in the number of patent filings by IP5 definition, worldwide number of filings



Source: EPO – PATSTAT; Fraunhofer ISI calculations.

When comparing the filing figures across the different IP5 definitions (**Figure 6**), it becomes visible that IP5 definition 1 has the largest number of patent filings by far. This makes sense, since single filings at the respective offices are included in this definition. IP5 definition 2 and definition 3 have similar levels, although the figures are a bit higher for definition 2. Interestingly, the trends differ between definition 1 and definition 2/3, especially in recent years. This becomes clearer when looking at **Figure 7**, where the country shares in worldwide patent filings are provided differentiated by the patent office definition. As we can see, Germany has its largest share in worldwide filings based on the transnational definition. Its shares are lower for all three IP5 definitions, especially definition 1. A similar picture can be observed for the US. The Asian countries, on the other hand, have larger shares in worldwide filings according to the IP5 definitions. This is because three Asian offices are included in the IP5 definition, which gives them a larger weight as they have a home advantage at their respective national and neighbouring offices.

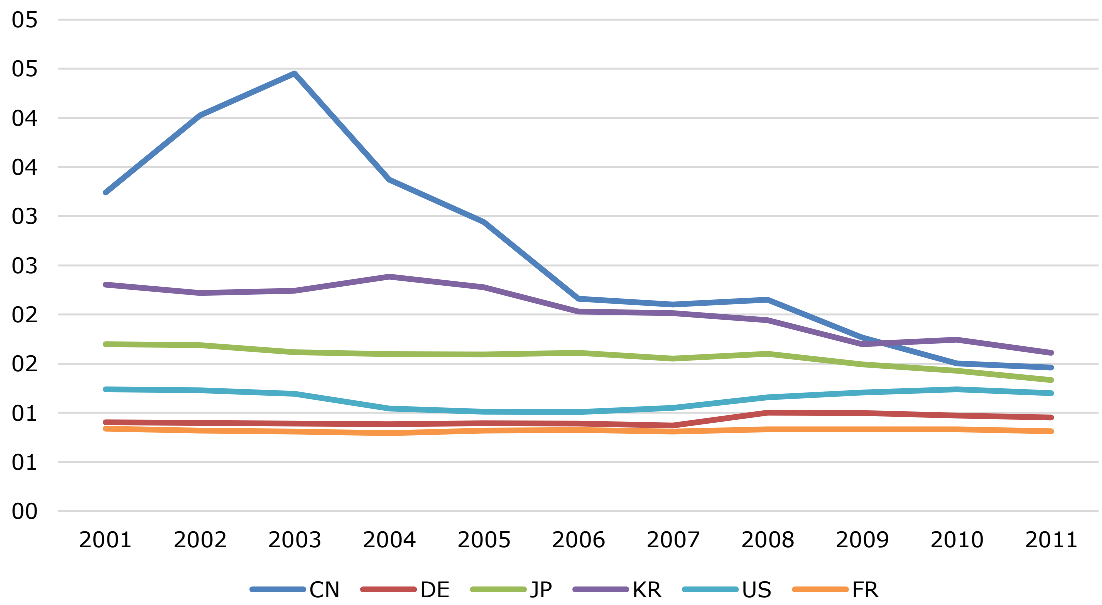
Figure 7: Country share in worldwide patent filings by patent office definition (Transnational vs. IP5, 2010-2012)



Source: EPO – PATSTAT; Fraunhofer ISI calculations.

This is especially visible for China according to definition 1, which gives a very large weight to single filings at SIPO, massively influencing the shares. This also becomes visible in **Figure 8**, where IP5 filings (definition 2) were divided by transnational filings, i.e. values above one indicate that the shares for country X are larger when using the IP-definition than using the transnational approach and vice versa. The trends are similar to those in **Figure 7**. The European countries Germany and France are scaled down, while the Asian countries are scaled up. The US remains more or less unchanged.

Figure 8: IP5 (Definition 2) filings divided by transnational filings, worldwide

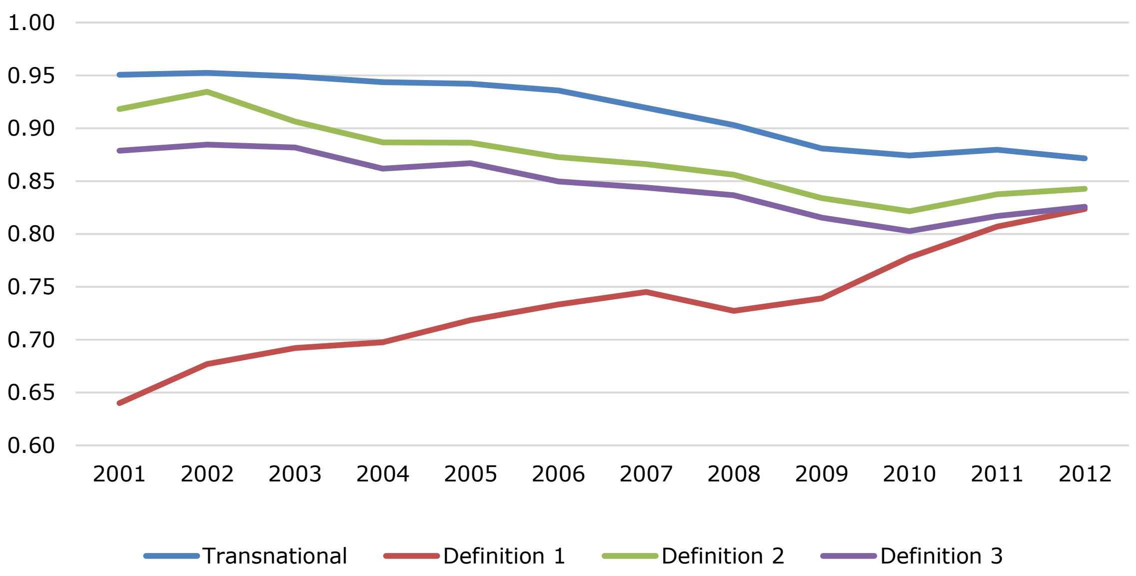


Source: EPO – PATSTAT; Fraunhofer ISI calculations.

In sum, it has to be kept in mind that the IP5-definitions favour Asian applicants oriented towards domestic and neighbouring markets. One proposal to offset this effect would be to generate a fourth IP5 definition which includes a filing at the EPO and/or the USPTO and one of the Asian patent offices, while the filing at the Asian offices should only be counted once, regardless of whether the filing targeted only one, two or three Asian offices. This would be more restrictive than the current definition 2, but less restrictive than the current definition 3 and it would counterbalance the large weight given to the Asian offices. However, we did not do this in the current project in order to be consistent with earlier statistics.

In a final analysis, we correlated the transnational approach and the different IP5 definitions with annual data on business R&D expenditures (BERD) obtained from the OECD. We found a high correlation between business R&D expenditures and patents, as expected, although the correlation has decreased slightly over the years. It can further be seen that the correlation is highest between BERD and patent filings counted using the transnational approach. The correlation is slightly lower for IP5 definitions 2 and 3. IP5 definition 1 is a special case as the correlation was very low at the beginning of the 2000s, which probably has to do with a massive rise in filings from Chinese applicants.

Figure 9: Pairwise correlation between IP5/transnational filings and BERD annual values, worldwide



Source: EPO – PATSTAT, OECD. Note: All correlations are significant at the $p<0.01$ level.

Overall, it can be stated that the definition of the patent office massively influences the results in terms of country rankings. This implies that the definition of the patent office should be chosen carefully depending on the aim of the analysis. When comparing the technological strengths and weaknesses of countries, the transnational approach seems most well-suited. It is able to overcome home advantages and unequal market orientations of the respective actors. Furthermore, it allows the most current analyses, as PCT filings are included, and it has the highest correlation to innovation inputs in terms of R&D. These can be seen as two further main advantages of the transnational approach.

Looking at single national (or regional) offices, on the other hand, makes most sense when focusing on the activities of specific countries on the respective market, i.e. when analysing market orientations towards the U.S. or China, it is most suitable to look at USPTO or SIPO filings only. It has to be kept in mind, however, that there is a home advantage of domestic applicants that influences the country rankings, which is always a problem when single filings (filings at only one patent office) are included.

With regard to IP5, three definitions have been proposed by the OECD and the European Commission. However, all three definitions seem to favour Asian applicants who naturally have an orientation towards their national and neighbouring markets. To balance the large weight currently given to Asian offices by the IP5 definitions, a fourth IP5 definition would be useful, where filings at the Asian offices are only counted once.

3 Descriptive data analyses

This section provides a descriptive overview of the data produced for each KET and each SGC. It does so by showing developments of research expenditures, patenting, value added, and employment over time within KETs and SGCs. It then benchmarks the structures and trends in SGCs and KETs in Europe against the trends in the USA and Asia. A comparative analysis is provided indicating how Europe compares to other countries worldwide in terms of research expenditures and patenting activities in KETs and SGCs.

3.1 Trends in KETs in the EU-28

3.1.1 Overall trends

The size and growth trend for the six Key Enabling Technologies (KETs) is described in terms of business R&D expenditures (BERD), employment, value added and patents (see **Figure 10** to **Figure 13**, respectively). Figures are in absolute terms and represent estimates by KET as a result of the application of the concordance scheme on the imputed database (see chapter 2). The data is subsequently used to calculate R&D shares and patent intensities as well as patent-to-R&D ratios.

For 2005-2012 the highest **R&D expenditure** in absolute terms can be found in Advanced Manufacturing Technologies (AMT). Compared to the second KET – Nanotechnologies - BERD within AMT was almost twice as large. The lowest R&D expenditure was observed in micro and nano-electronics, biotechnology and photonics. Within the period there was a moderately positive trend: all six KETs followed a similar growth pattern of about 4.6% average annual growth (AAG)[10].

Between 2005 and 2014, the highest **employment figures were observed** in Advanced Manufacturing. Employment was almost double the second KET, Advanced materials. The lowest employment was seen in micro and nano-electronic technologies. Average annual employment growth has remained essentially stable over the period for all KETs with about 0.2%.

In terms of **value added,** AMT ranked top and was four times larger than advanced materials in second place. The lowest value added occurred in micro and nano-electronic technologies. The trend in value added was quite stable with 1.3% on average annually, though, two notable moments of decline were observed in 2009 and 2014. The biggest decline in 2014 was in advanced materials (ca. -17%).

Finally, the volume of **transnational patents** was the highest in advanced manufacturing followed by advanced materials. The gap between AMT and the other KETs was not as pronounced for patents as it was in terms of BERD, employment and value added. Nanotechnologies, while second or third highest in BERD, employment and value added ranked last in terms of patents and has also experienced the steepest decline since 2011 (ca. -72%).[11] On the other side, the KET Micro and nano-electronic technologies, while ranking last in BERD, employment and value added, positioned third

[10] The average annual growth rate is calculated according to the following formula:

$$AAG = (((EV/SV)^{(1/n)}) - 1)*100$$

where EV is the ending value, SV is the starting value, and n is the number of years between the two values. Since only the starting and ending value are taken into account for this calculation, specific trends within the given time period, e.g. the financial crisis in 2008/2009, are not taken into account by this indicator. It further has to be kept in mind that the data availability, especially in the recent years, differs across indicators, implying that the starting and end values are not the same across indicators. The alternative would have been to use the same time window for all indicators, which, however, would have resulted in a loss of information for the indicators for which most recent data is available.

[11] The PATSTAT version from September 2016 results in decreasing numbers of patent filings in nanotechnology. This holds for IPC and CPC-based searches. However, a keyword-based search in World Patent Index in May 2017 shows a strictly increasing trend.

in terms of patenting. But also here, a declining trend was seen since 2011 (ca. -14%). The average KET growth in patenting was rather stable with a marginal decline (-0.1% on average annually).

Figure 10: BERD (in EUR billion)

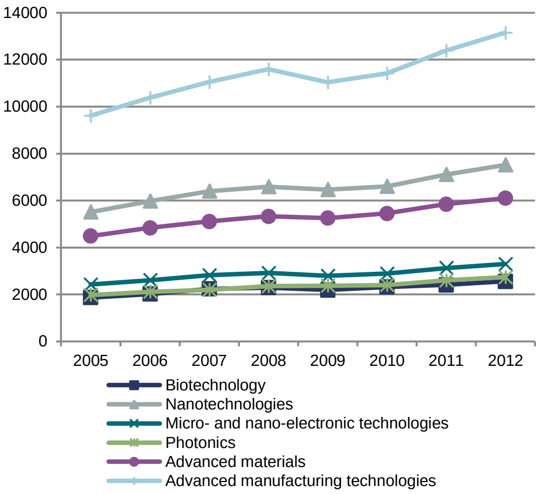


Figure 11: Employment (in million persons)

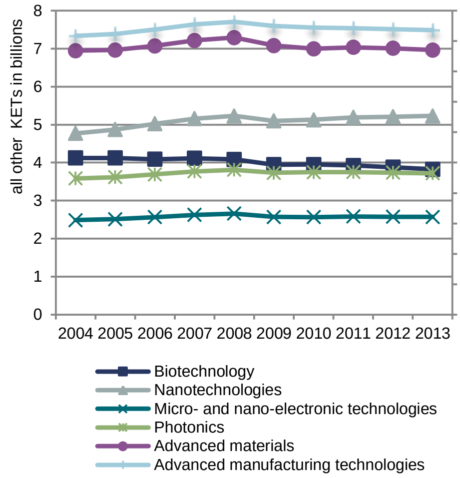


Figure 12: Value Added (in EUR billion)

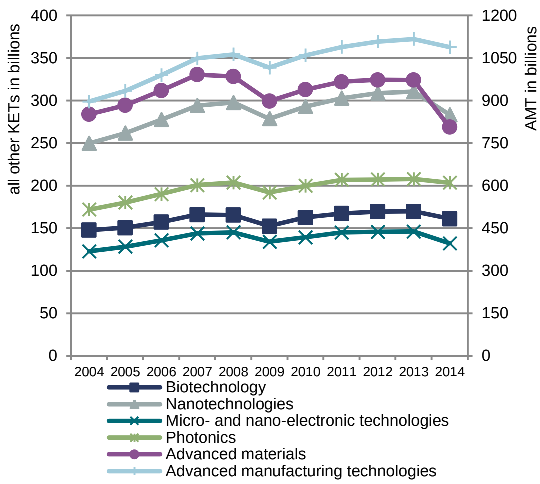
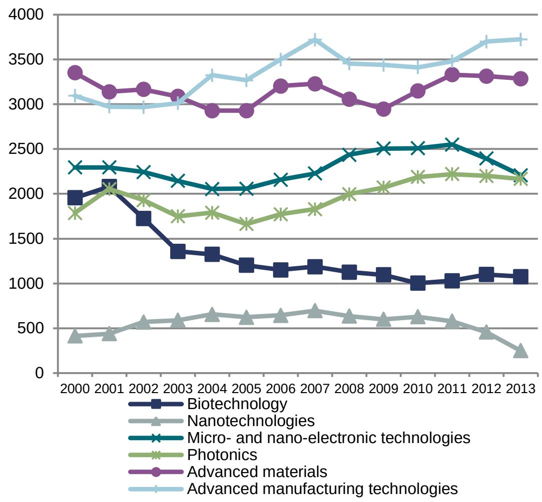


Figure 13: Patents (counts)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Notes: The reference period varies by indicator due to the respective data coverage; Note that for Main Economic Indicators (Value Added and Employment) the right axis is used to represent the sector with the highest value, i.e. Advanced Manufacturing Technologies. BERD as a percentage of GDP and separate indicators for each KET are presented in Annex.

3.1.2 Profile of KETs

The characteristics by KET in EU28 can be described as follows:

— **Biotechnology**, with an R&D expenditure of ca. EUR 2.5 billion in 2012, has been in relative terms, i.e. compared to other KETs, among those with the lowest expenditures in R&D throughout the entire period (2005-2012). The annual average growth rate in R&D expenditure over 2005-2012 is similar to the other KETs (4.6%).

With ca. 3.9 million employees in 2014, Biotechnology ranked fourth among the six KETs. Employment growth has been slightly negative since 2004 (-0.6% aag) rendering Biotechnology the only KET experiencing a negative employment trend. In terms of value added (EUR 161 billion in 2014) Biotechnology scored only fifth out of the six KETs. While recording positive average annual growth of 0.87% over the period 2004-2014, it experienced a sudden decline of 8% between 2008 and 2009, but quickly recovered in 2010. **In terms of patenting**, Biotech experienced its peak in 2001 followed by a sharp decline in 2002 and 2003 and stability over the last ten years 2003-2013.

— R&D expenditure in **nanotechnologies** was about 7.5 EUR billion in 2012, ranking this KET second over the 2005-2012 period. Average annual growth rate since 2005 was 4.6%, which is comparable to other KETs. Nanotechnologies represent the third largest KET in terms of employment, with about 5.2 million employees in 2014. Employment growth since 2004 is the highest within KETs (average annual growth of 0.9%). The value added associated with Nanotechnologies is EUR 283 billion, which is the second largest figure amongst the six KETs. Growth rate of value added has also been the second highest since 2004 (1.3% aag), even though a decrease of 6% was observed over 2008-2009. This KET presents the lowest figures in terms of patents, with a slight decline over the most recent years.

— **Photonics** had an R&D expenditure of EUR 2.7 billion in 2012. Compared to other KETs, it was among those with the lowest expenditures in R&D throughout the period 2005-2012. However, in terms of average annual growth it ranked among the highest with 4.7%. In terms of employment (3.7 million employed in 2014), Photonics ranked fifth among the six KETs. Employment growth has been stable with a slight increase of 0.3% on average annually. The value added (EUR 203 billion in 2014) positioned Photonics fourth among the six KETs. While there was a positive growth of 1.7% aag between 2004-2014, it experienced a drop of 6% between 2008 and 2009 which was quickly recovered in 2010. Finally, with 2,167 patents, it ranked fourth among the KETs and has experienced an average annual growth of 1.5% over the period 2000-2013.

— Compared to other KETs, **Micro and Nano-electronic technology** was with an R&D expenditure of ca. EUR 3.3 billion in 2012 also among those with the lowest expenditures in R&D throughout the entire period (2005-2012). In terms of average annual growth (4.5%) it followed the average growth for all KETs. An employment rate of 2.6 million persons in 2014 ranked Micro and Nano-electronic technology fourth among the six KETs. Employment growth remained stable with a slight increase of 0.4%. The value added of ca. EUR 132 billion in 2014 positioned Micro and Nano-electronic technology fourth among the six KETs. Like the other KETs it recorded a positive growth over the period 2004-2014 (0.7%), but it also experienced a decline of 8% between 2008 and 2009 with a quick recovery in 2010. When it comes to patents, it ranked third among the six KETs with 2,207 patents but experienced a slight decline of 0.3% on average annually over the period 2000-2013.

— With a R&D expenditure of ca. EUR 6.1 billion in 2012, **advanced materials** ranked third throughout the entire period (2005-2012). The growth rate in R&D expenditure was 4.5% following the average growth for all KETs. The employment of 6.9 million persons in 2014 ranked advanced materials second among the six KETs. The employment growth has been stable with a marginal increase of 0.06%. The value added of ca. EUR 269 billion in 2014 positioned advanced materials third among the six KETs. A negative average annual growth of -0.5% was seen over the period 2004-2014, with two notable drops one in 2009 with a decline of 9% between 2008 and 2009 and a quickly recovered in 2010 and the second in 2014 with a decline of 17%. Advanced materials was with 3,285 patents ranked second among the KETs and experienced a small decline of 0.16% per year over the period 2000-2013.

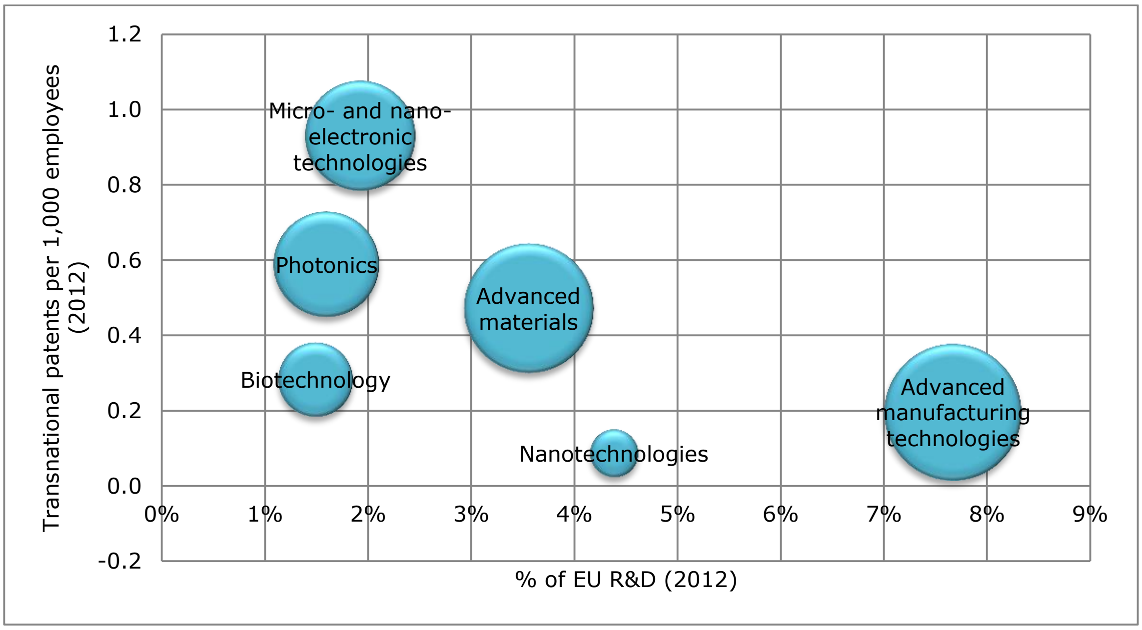
— **Advanced Manufacturing Technologies (AMT)**, with a R&D expenditure of ca. EUR 13.1 billion in 2012, ranked first throughout the entire period (2005-2012). The average annual growth rate in R&D expenditure of. 4.6% followed the average growth for all KETs. An employment rate with 18.6 million persons in 2014 ranked AMT first among the six KETs with a significant distance from the second largest Advanced materials. The employment growth has been stable with a minor increase of 0.16% on average annually. The value added of ca. EUR 1,088 billion in 2014 positioned Advanced Manufacturing Technologies first among the six KETs. The positive annual average growth of 1.9% was also the highest among all KETs over the period 2004-2014. Equal to the other KETs, a decrease of 4% was seen between 2008 and 2009 - the lowest among the KETs, and a quick recovery happened in 2010. In terms of patents, AMT ranked first with 3,725 patents and has experienced an annual increase of 1.44% over the period 2000-2013.

R&D shares and patent intensities

Figure 14 positions each KET in terms of share of business R&D over the total EU BERD and patent intensity. This data was estimated based on the matrices and methods described in section 2.3.This graph illustrates the extent to which KETs are important in the total R&D budget (horizontal axis) and relates the R&D figures to its output in terms of patent filings per employee (vertical axis). The size of the circles indicates the size of the technology in terms of number of transnational patents. The data refers to 2012.

The share of business R&D is the highest for **advanced manufacturing technologies** (7.7%). Nonetheless, the same KET also registers a low patent intensity (0.2), calculated as the number of transnational patents per 1,000 employees. The second highest share of business R&D has been registered for **nanotechnologies** (4.4%). Similarly to Advanced Manufacturing Technologies, **nanotechnologies** show the lowest patent intensity (0.1). Nanotechnologies, also register the lowest absolute number of patents among the KETs (458). **Advanced materials**, second only to AMT in terms of size, display a moderate share of business R&D (3.6%) and patent intensity (0.5). Finally, **biotechnology, photonics** and **micro-and nano-electronic technologies** have similar R&D shares (1.5%, 1.6% and 1.9% respectively), but perform differently in patent intensity; while **biotechnology** (0.3) is in line with the average patent intensity (0.4), patent intensity is higher than the average for **photonics** (0.6) and the highest in relative terms for **micro-and nano-electronic technologies** (0.9).

Figure 14: R&D shares and patents per 1,000 employees for KETs (size of circles: number of transnational patents) - 2012



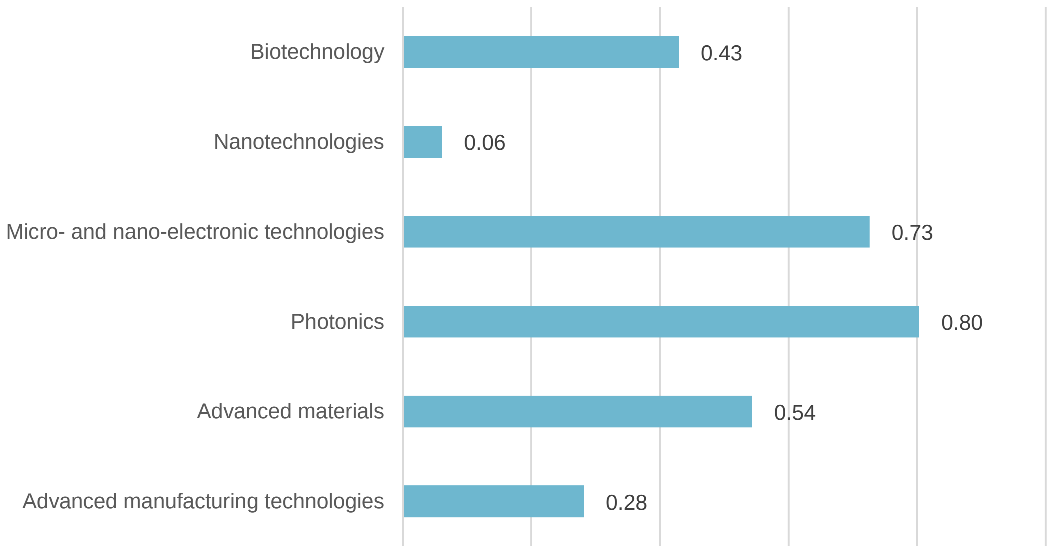
Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Patent-to-R&D ratio

Figure 15 presents the number of transnational patent applications per million EUR of R&D expenditures (patent-to-R&D ratio or propensity-to-patent). This ratio can be interpreted to some extent as an indicator of research productivity (with patents as an output produced from research activities). However, strong variations can also be explained by differences in patenting propensities between industries and fields that reflect strategic propensities and appropriability propensities (de Rassenfosse and van Pottelsberghe, 2009).

Figure 15 shows that **Photonics** (0.80), **micro- and nano-technologies** (0.73) and **advanced materials** (0.54) have the highest patent-to-R&D ratios in the EU28. **Biotechnology** (0.43) and **AMT** (0.29) show a moderate number of patents per EUR spent. **Nanotechnologies,** with a significantly lower patent-to-R&D ratio (0.06) ranks last. Some contextual information is provided by the KETs Observatory (2015)[12] pointing out that KETs results interpretation should keep in mind that **photonics** and **micro- and nano-electronic technologies** are also leading sectors in terms of trade performance, while **nanotechnologies** and **biotechnolog**y are more associated with production processes (and less with products), thus accounting for a smaller share of the EU KETs market overall.

Figure 15: Patent-to-R&D - KETs (transnational patents per EUR million of R&D)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

3.1.3 Comparative analysis

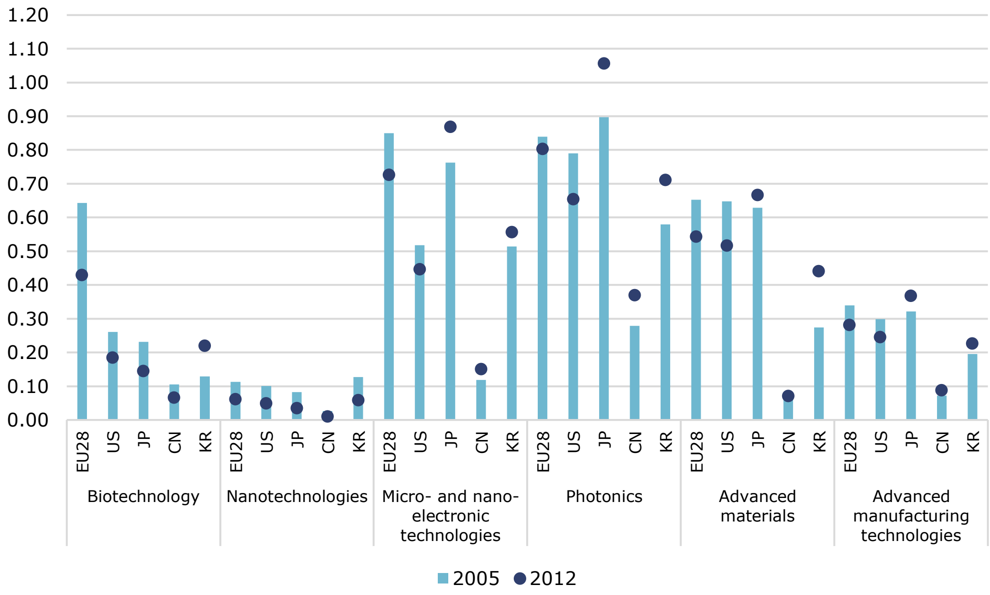
The analysis performed for the EU28 has been replicated for the US, Japan, China and South Korea with the purpose of comparing the EU28 performance to the performance of other major countries. The outcome is summarised in **Figure 16**, comparing the patent-to-R&D ratio (reflecting research productivity and patenting propensity) between the countries in two different points in time (2005 and 2012) across the six KETs (graphs for BERD and Patents are included in the Annex).

According to the data estimated for 2012 the EU28 ranked first in terms of **patents per million EUR of R&D** in biotechnology. In micro and nano-electronic technologies, photonics, advanced materials and advanced manufacturing technologies it ranked

[12] EC (2015), KETs Observatory, p.19m, https://ec.europa.eu/growth/tools-databases/kets-tools/sites/default/files/library/kets_1st_annual_report.pdf

second after Japan. In nanotechnology, all countries besides China demonstrated a similar patent-to-R&D ratio. In 2005, the EU28 ranked first in biotechnology and AMT. In the remaining nanotechnologies, micro and nano-electronic technologies, and advanced materials, it ranked third after the changing leaders US or Japan.

Figure 16: International comparison of Patent-to-R&D (transnational patents per EUR million of R&D) - KETs



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

— In **biotechnology**, the patent-to-R&D ratio in 2012 was the highest in the EU28 followed by South Korea, the US, Japan and China. The EU28 had experienced a sharp decline of 33% in the patent-to-R&D ratio in 2012 compared to 2005 which was driven by a decline in patents (ca. 9%). South Korea is the only country that experienced a positive growth in this ratio in 2012 closing the gap with the US and Japan and even exceeding the latter countries.

— In their patent-to-R&D ratio for 2012 for **nanotechnology,** the EU28, South Korea, the US and Japan were rather similar (0.04 Japan, 0.05 the US, 0.06. EU28 and South Korea). In 2005, South Korea had been leading (0.13) while the EU28 was positioned between the US and Japan (0.11). By 2012, all countries had experienced significant declines in their patent-to-R&D ratio. The EU28 encountered the smallest (but still 46%) and China the highest (62%). All countries increased their BERD in this period but patents did not follow at a corresponding pace.

— In 2005, the EU28 ranked top in terms of patent-to-R&D ratio in **micro and nano-electronic technologies** (0.85), followed by Japan (0.76) and the US and South Korea (0.50). In 2012 however the EU28 lost its leading position to Japan. The EU28 and the US were the only countries experiencing a decline of -14% each while all others increased their patent-to-R&D ratio, particularly China by. 27%.

— In 2012 Japan had the highest patent-to-R&D ratio in **photonics**, (1.06), followed by the EU28 (0.80), South Korea (0.71) and the US (0.65). Compared to 2005, the EU28 and the US were the only countries experiencing declines in their patent-to-R&D ratio (-17% in the US and -4% in the EU28). China increased its BERD five times while South Korea doubled its BERD. In terms patents, Japan excelled with an increase of 49%.

— In **Advanced materials**, Japan had the highest patent-to-R&D ratio in 2012 followed closely by the EU28 and the US. These two had decreased their patent-to-R&D ratio by 17% and 20% respectively compared to 2005. While BERD increased in both countries – in particular in the EU28 by 38% - patents grew also in the EU28 (13%) while they decreased in the US (-13%). As in the case of photonics, China increased its BERD five times and South Korea doubled its BERD. In terms of patents, China experienced almost a fourfold increase, South Korea a threefold increase and Japan an increase of ca. 45%.

— In **Advanced Manufacturing Technologies**, Japan led in 2012 followed by the EU28 and the US and then South Korea and China. The EU28 BERD and patents were significantly higher compared to all other countries (66% more BERD and 60% more patents than the second ranked US (2012)). Other countries increasing their investments significantly such as China and South Korea with a fourfold and twofold increase respectively. In terms of patents, China saw a more than six fold increases, South Korea doubled its patents while Japan also increased its patent numbers by 43%.

Specialisation (Revealed Technology Advantage, RTA) for the EU28 can be calculated by observing the relative share of BERD or patents for each KET in comparison with other countries. The RTA is one of the most prominent indicators of technological specialisation, that has been described in (Grupp 1998). The equation for the indicator is inspired by Balassa's RCA and can be represented as follows:

$$RTA_{ij} = 100 \cdot \text{tanhyp}\left(\log \frac{p_{ij} / \sum_{k=1}^{J} p_{ik}}{\sum_{h=1}^{I} p_{hj} / \sum_{h=1}^{I} \sum_{k=1}^{J} p_{hk}} \right)$$

The RTA for country i in technology j measures the share of BERD/patents of country i in technology j compared to the world share of BERD/patents in technology j. If a country i's share is larger than the world share country i is said to be specialised in this technology. The tanhyp-log transformation does not change this general interpretation but it symmetrises this indicator by normalising to an interval ranging from -100 to +100.

Specialisation indices were produced for the period 2010-2012 based on the relative weight of each KET in Europe compared to the weight of each KET in all countries (i.e. all countries available in the dataset henceforth labelled as "world"). Large positive (resp. negative) values illustrate high (low) specialisation in the technology.

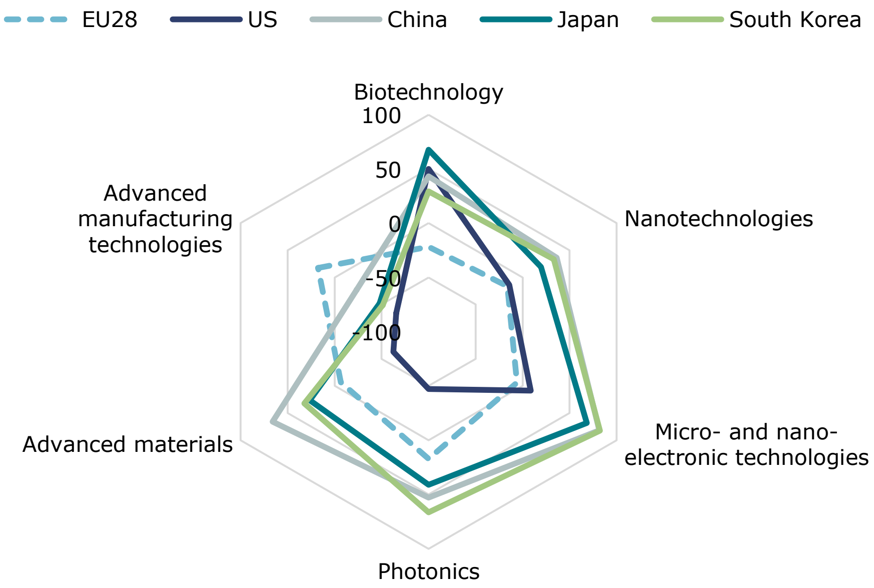
BERD based indices show that the EU28 was around field average for all KETs with the exception of biotechnology (index: -22) in which it demonstrated a slightly lower specialisation compared to the "world". Patent based indices showed that the EU28 was around field average for all KETs besides micro and nano electronic technologies (index: -48) and biotechnology (index: -23) in which the EU28 demonstrated a lower specialisation compared to the "world". In Advanced Manufacturing Technologies, the EU28 had a higher specialisation compared to the "world" (index: 29).

Figure 17 and **Figure 18** illustrate the specialisation indexes for EU28 KETs in terms of R&D (BERD) and transnational patents (TNL) calculated for the EU28, the US, China, Japan, and South Korea. The EU28 had the highest specialisation for advanced manufacturing technologies in terms of both R&D and (especially) patents.

The US showed a strong specialisation in biotechnology both in terms of R&D and patents (specialisation indices being respectively 50 and 38), and only based on patents in nanotechnologies (index: 28). The specialisation pattern in terms of R&D was similar between China, Japan and South Korea: the three countries had a very high specialisation in R&D related to micro- and nano-electronic technologies and to a lesser extent photonics, advanced materials (especially China), biotechnology (especially Japan), and nanotechnologies.

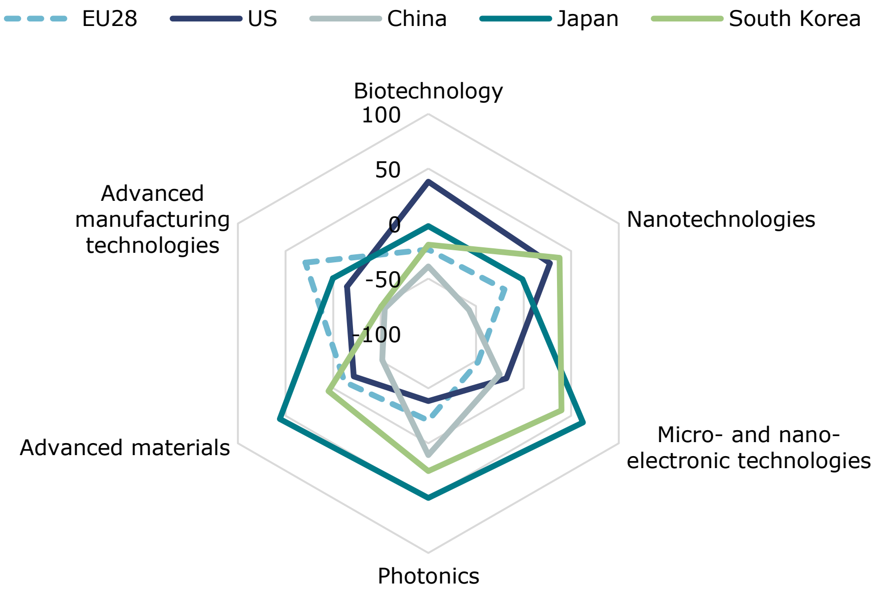
Regarding specialisation in terms of transnational patents, China showed lower shares of patents in all KETs except photonics compared to the world. Japan patented much more than other countries in photonics (index: 50), advanced materials (index: 56) and micro- and nano-electronic technologies (index: 62). South Korea showed a patent specialisation in nanotechnologies (index: 38), micro- and nano-electronic technologies (index: 40), and photonics (index: 25).

Figure 17: Specialisation EU28 vs. selected countries (based on R&D expenditure)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Figure 18: Specialisation of EU28 vs. selected countries (based on transnational patents)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

3.2 Trends in SGCs in the EU-28

3.2.1 Overall trends

The size and growth trend for the six Societal Grand Challenges (SGCs) are described in terms of business R&D expenditures (BERD), employment, value added and patents (see **Figure 19**, **Figure 20**, **Figure 21**, and **Figure 22** respectively). The figures are in absolute terms and are estimated per SGC - the result of the application of the concordance scheme applied on the imputed database (see chapter 2.3). The latter data are subsequently used to calculate R&D shares and patent intensities as well as patent-to-R&D ratios.

The highest **R&D expenditure** in absolute terms throughout the entire period (2005-2012) was in Transport. This was more than five times higher than the second SGC, Health (in 2012). The lowest R&D expenditure occurred in Climate. The growth in BERD across the six SGCs followed a similar pattern with ca. 4.7% annual average growth.

In terms of Employment (2005-2014)[13] Transport had the highest numbers and was twice as large as Health, the second largest SGC (in 2014). The lowest employment was in Climate. Employment growth has remained essentially stable over the period for most SGCs (0.2% average annual growth). Exceptions were Health (1%) and Climate (-0.7%).

In terms of **value added,** Transport ranked at the top throughout the entire period (2005-2014) and enjoyed a more than two times higher VA than the second ranked Health (in 2014). The lowest value added was recorded for Climate. The trend in value added was moderately positive (1.7%) for the whole period with two notable exceptions, in 2009 and 2014. The decline in 2014 was however less pronounced and experienced only in food, agriculture, bioeconomy, climate (each -2%) as well as energy and transport (each -7%).

For the volume of **patents** throughout the entire period 2000-2013, it was the highest for transport followed by food, agriculture, bioeconomy and then health. The gap between transport and the following two SGCs was not as pronounced for patents compared to BERD, employment, and value added. The average SGC growth in patenting remained stable with a slight average annual growth of 0.4%. The most volatile pattern registered was in the SGC energy with a growth of 15% in 2004-2011, followed by a sharp decline in 2013 (ca.60%).

[13] The difference in the periodic coverage stems from the fact that always the latest available data was used.

Figure 19: BERD (in EUR billion)

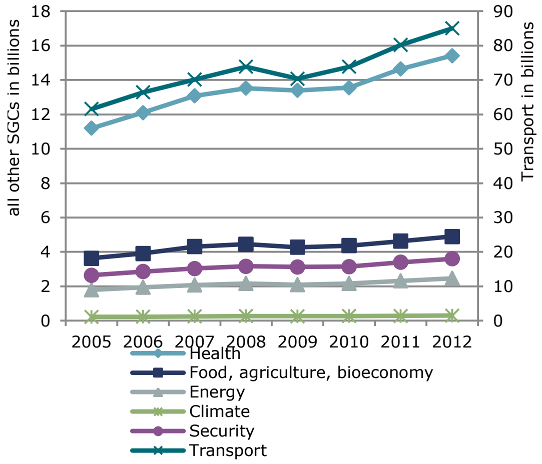


Figure 20: Employment (in million persons)

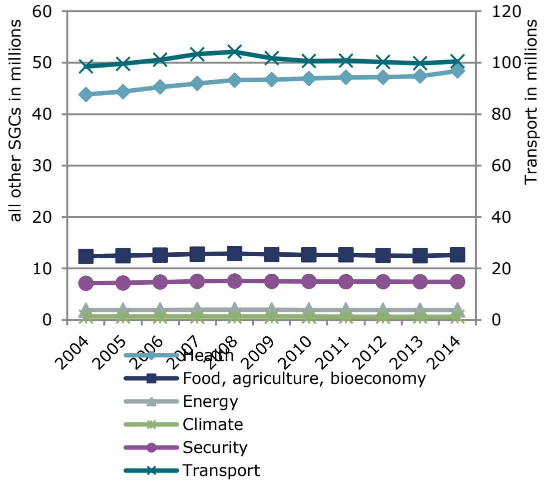


Figure 21: Value Added (in EUR billion)

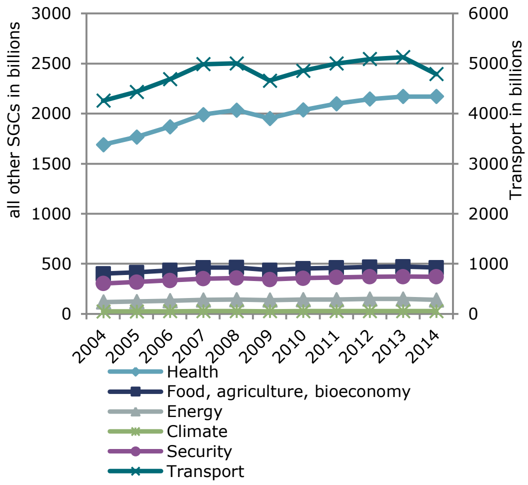
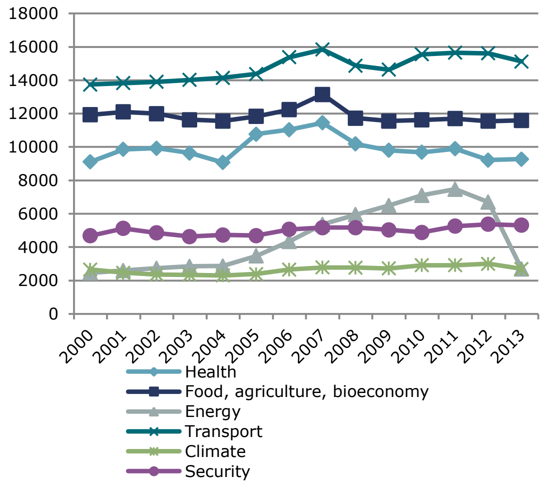


Figure 22: Patents (in counts)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Notes: The reference period varies by indicator due to the respective data coverage; Note that for BERD and Main Economic Indicators (value added and employment) the right axis is used to represent the sector with the highest values. BERD as a percentage of GDP and separate indicators for each SGC are presented in Annex.

3.2.2 Profile of SGCs

Observations made by SGC are included in the paragraphs below.

— Compared to other SGCs **Climate** was with an R&D expenditure of EUR 0.3 billion in 2012, among those with the lowest expenditures in R&D throughout the entire period (2005-2012).[14] The average annual growth of 4.5% was similar to the other SGCs. An employment share of 0.6 million persons in 2014 ranked Climate last among the six SGCs. The employment growth decreased moderately by -0.7%. Climate is thus the only SGC with a negative employment trend. The value added of about EUR 27 billion in 2014 was the lowest among the SGCs. There was even a significant distance to the second last SGC (Energy with EUR 139 billion). While recording a positive growth of 1.36% on average annually in VA over the period 2004-2014, it

[14] The different time periods mentioned in these sections result from data availability. We always used the latest available data. Our time series usually started in 2006 for the economic variables (due to NACE Rev 2 coverage) and in 2000 for of the patent data.

experienced the sharpest decline among SGCs between 2008 and 2009 (-8%) but quickly recovered in 2010. Finally, Climate had the lowest volume of patents among the SGCs with a grand total of 2,699 in 2013 and the lowest average annual growth of 0.1% over the period 2000-2013.

— **Compared to the other SGCs,** the R&D expenditure of EUR 2.5 billion in (2012) in **Energy** was among the lowest expenditures throughout the entire period (2005-2012). The average annual growth in R&D expenditure was similar to the other SGCs (4.6%). In terms of employment, 1.9 million persons in 2014 made Energy rank second last among the six SGCs. The employment growth has been stable with a slight increase of 0.1%. Also in terms of value added - EUR 139 billion in 2014 - positioned Energy second last in this category. However, the drop of 2% between 2008 and 2009 was the lowest among the SGCs which it quickly recovered in 2010. It encountered a second drop between 2013 and 2014 and here the decline was the highest with 7% (together with Transport).[15] Finally, Energy had the second lowest volume of patents among the SGCs, a grand total of 2,701 in 2013 and the third highest growth of ca. 0.73% over the period 2000-2013.

— **Security**, with an R&D expenditure of ca. EUR 3.6 billion in 2012, was in relative terms, among those SGCs with the lowest expenditures in R&D throughout the entire period (2005-2012). The average annual growth was similar to the other SGCs with 4.5%. Employment of 7.4 million persons in 2014 ranked Security forth among the six SGCs. The employment growth was stable with a slight increase of 0.4% over the whole period. The value added of EUR 372 billion in 2014 positioned Security forth among the SGCs. Overall, it recorded a positive growth of 1.4% over the period 2004-2014 and with 2% it experienced the smallest decline in 2009 among all SGCs. Security had the third lowest volume of patents among the SGCs a grand total of 5,312 in 2013 and the highest average annual growth of 1% over the period 2000-2013.

— **Food, agriculture, bioeconomy**, was with an R&D expenditure of EUR 4.9 billion in 2012, the third largest SGC in this category throughout the entire period (2005-2012) but with a significant distance from the first two SGCs. The average annual growth in R&D expenditure was similar to the other SGCs with 4.4%. In terms of employment, with 12.6 million persons in 2014, it ranked third among the six SGCs. The average annual employment growth remained rather stable with a slight increase of 0.2%. The value added of EUR 462 billion in 2014 positioned Food, agriculture, bioeconomy third among the SGCs. While it showed a positive growth of 1.4% over the period 2004-2014, it experienced a decline in 2009 similar to the other SGCs (6%), which it quickly recovered in 2010. Finally, Food, agriculture, bioeconomy had the second highest volume of patents among the SGCs (even higher than Health which ranked second in all other indicators expect patents) with a grand total of 11,596 in 2013 slight decline of 0.2% on average annually over the period 2000-2013.

— **Health**, with an R&D expenditure of EUR 15.4 billion in 2012, was compared to the other SGCs the second largest in R&D expenditures throughout the entire period (2005-2012). With 4.7% average annual growth, R&D expenditure increased slightly more than the average of the SGCs. The employment of 48.4 million persons in 2014 ranked it second among the six SGCs. The average annual employment growth of 1% was the highest among the SGCs. In terms of value added (EUR 2,171 billion in 2014) Health comes second with both large distances from the leading Transport and the third Food, agriculture, bioeconomy. Health enjoyed the highest positive growth in VA of 2.5% over the period 2004-2014, but it also experienced a decline of 4% in 2009 which it quickly recovered in 2010. Health has the third highest volume of

[15] This decline in energy has a methodological reason, i.e. it is only due to the classification of the energy SGC, namely the Y-classification. A disadvantage of the Y-classification is that the CPC (Cooperative Patent Classification), on which it is based, is not provided for patents until the patents pending via the PCT process are transferred to the national phase. This is the case only 30 months after registration. The current margin of the figures is therefore even further back than in purely IPC based patent searches.

patents among the SGCs with a grand total of 9,277 in 2013 and a stable growth with only a slight increase of 0.1% over the period 2000-2013.

— With a R&D expenditure of EUR 85 billion in 2012, **Transport** was in relative terms by far the largest SGC in this category throughout the entire period (2005-2012). The average annual growth in R&D expenditure was slightly higher than the average KET with 4.7%. The employment of 100.5 million persons in 2014 ranked it first among the six SGCs - this is twice as high as employment in Health. The slight employment growth of 0.2% was equal to the average of the six SGCs. In terms of value added (EUR 4,789 billion in 2014) was more than double the VA of Health – ranking second highest among the SGCs. While Transport recorded a positive growth of 1.2% on average annual for VA over the period 2004-2014, it also experienced a decline of 7% in 2009 – from which it quickly recovered in 2010/2011. Finally, Transport had the highest volume of patents among the SGCs with a grand total of 15,124 in 2013 and a growth of 0.7%, the second highest after Security over the period 2000-2013.

R&D shares and patent intensities

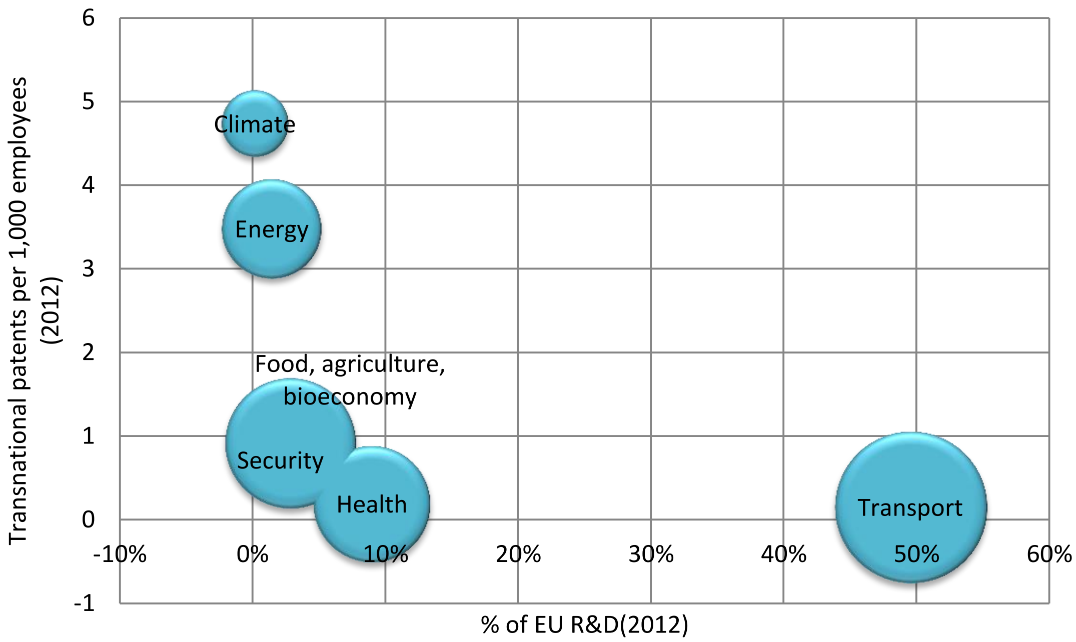
Figure 23 illustrates the performance of SGCs in terms of share of total EU R&D and patent intensity. This figure shows the extent to which SGCs amount to important levels of R&D expenditure in total EU R&D and it also rescales patents with respect to total employees in the SGC. The size of the bubble indicates the size of the SGCs in terms of number of transnational patents. All data refers to 2012.

Transport showed the highest share of business R&D (49.5%). Despite registering the highest number of patents (15,615 in 2012), Transport was one of the least intensive SGCs in terms of patents per 1,000 employees. Similar to transport, also the **health** SGC had a very low patent intensity (0.2), and a non-negligible total number of patents (9,219 in 2012), ranking third after transport and the food, agriculture and bioeconomy SGC.

Food, agriculture and bioeconomy and **security** performed similarly in both R&D shares (2.9% and 2.1% respectively) and patent intensity (0.9 and 0.7 respectively). However, the **food, agriculture and bioeconomy** SGC counted more than twice the number of security patents.

Energy and **climate**, showed the highest patent intensity compared to the other SGCs (3.5 and 4.7 respectively). However, the total number of patents was the lowest across all SGCs (together with security). Energy was the second best performer in patent intensity and also counted more than twice the number of patents of climate (6,693 in 2012). While showing the highest patent intensities, energy and climate had at the same time the lowest business R&D shares (1.4% and 0.2%).

Figure 23: R&D shares and patents per 1,000 employees for SGCs (size of circles: number of transnational patents) - 2012



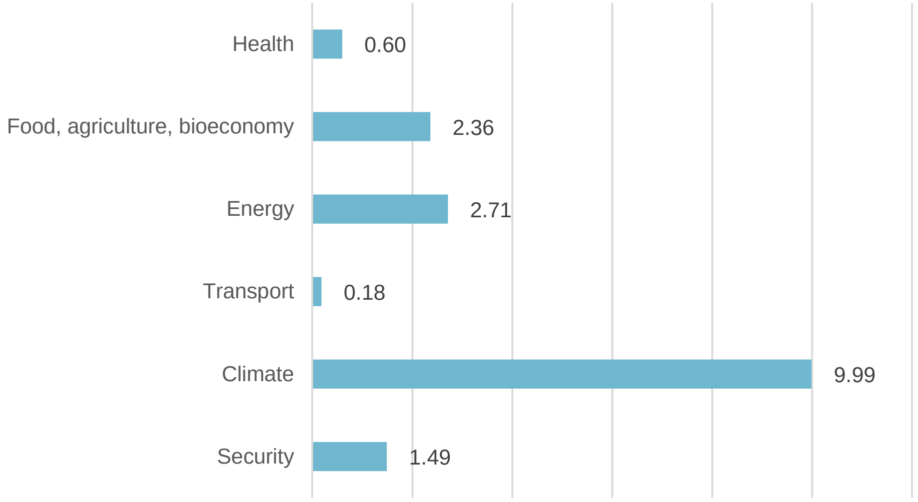
Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Patent-to-R&D ratio

Figure 24 presents the number of transnational patent applications per EUR million of R&D (patent-to-R&D ratio). This indicator can be interpreted as a measure that correlates with research productivity and patenting propensity related to appropriability and strategic factors. Hence, it shows how many patents can be related with 1 million euros spend in R&D for each SGC, but interpretation should be made in light of these different factors.

Figure 24 shows that the patent-to-R&D ratio in **climate** is significantly higher than in the rest of the SGCs (9.99). Conversely, an extremely low ratio is registered in **transport** (0.18). The two findings are in line with the trends on patent per employee. Similarly, the low number of patents per employee registered for **health** is mirrored by a low patent-to-R&D ratio (0.60). **Food, agriculture, bioeconomy** and **energy** show a moderate patent-to-R&D ratio, however not negligible (2.36 and 2.71 respectively). Compared to a very high patent intensity (expressed in transnational patents per 1,000 employees), the energy sector has a relatively lower number of patents per EUR spent.

Figure 24: Patent-to-R&D ratios for the six SGCs (transnational patents per EUR million of R&D)[16]



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

3.2.3 Comparative analysis

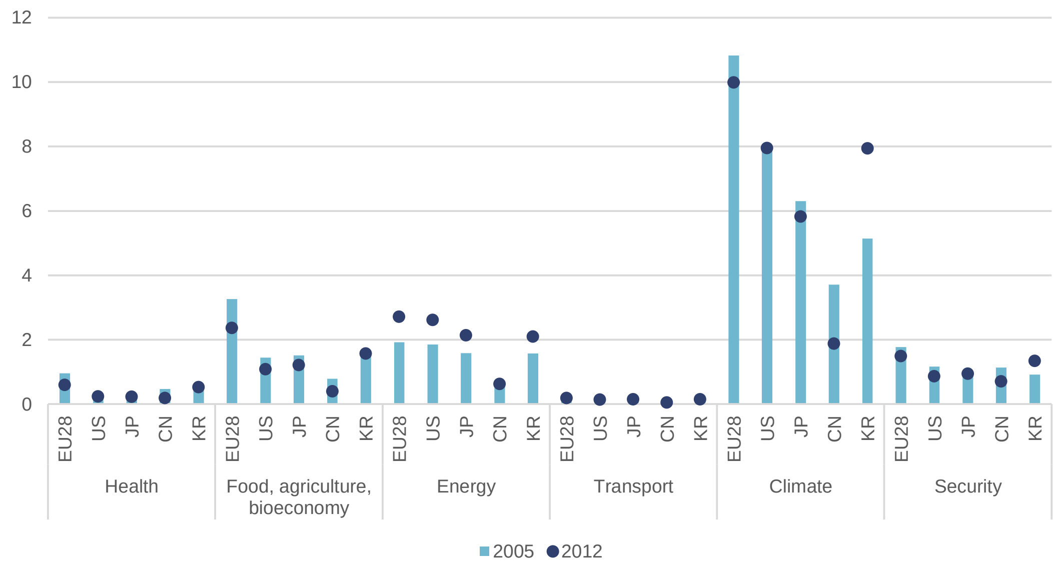
The analysis performed for the EU28 has been replicated for the US, Japan, China, and South Korea with the purpose of comparing the EU28 performance to the performance of other major countries. The outcome is presented by comparing the patent-to-R&D ratio between the countries in two different points in time (2005 and 2012) and across the six SGCs. The indicator used is the number of transnational patents per EUR million R&D (graphs for BERD and Patents are included in the Annex).[17]

According to the data estimated for 2012, the EU28 showed the highest patent-to-R&D ratio in all six SGCs. In terms of BERD the EU28 ranked first in energy, transport and climate. In the other SGCs health, food agriculture, bioeconomy and security, it ranked second after the US. In terms of patents the EU28 ranked first in all but health, where it ranked second to the US.

[16] The patent data for the SGCs are partially overlapping. The reason is that there are a few overlapping definitions, but the main overlap results from the fact that patents are assigned multiple IPC/CPC classes, of which some might fall into different categories of our definition. In addition, "Energy" was defined based on the EPO classification of climate change and mitigation technologies (Y-classes), which are a supplement to the IPC. Therefore, all patents with a Y-class also have "regular" IPC classes in addition, which might also result in overlaps, mainly with "Climate". For more details on this issue, please refer to the discussion paper (Frietsch et al. 2016).

[17] It may appear that we run into a kind of circular argumentation here as we use the patents to recalculate R&D by technologies and afterwards relate R&D and patents per technology field. This is not the case, however, mainly for two reasons. On the one hand, the R&D per technology does not only originate in one sector, but in a number of sectors. We use the sum of sector-specific R&D expenditures per technology. On the other hand, we use external weights to give each patent a different weight per sector. This results in a further decoupling of the patent the R&D data. For further explanations, please refer to the methodology section (see section 2.3).

Figure 25: International comparison of Patent-to-R&D (transnational patents per EUR million of R&D) - SGCs



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

— In **health**, the patent-to-R&D ratio as estimated for 2012 has been the highest in the EU28 (0.60) followed by South Korea (0.52) and the US and Japan (0.23 each). South Korea's second highest ratio is explained by its very low BERD (in relative terms). On the other hand, BERD of the US was more than three times as high as the EU28 BERD in 2012, while the number of patents was 30% higher than the EU28 equivalent. Also Japan has a higher BERD than the EU28 but less than half of the EU28 patents. In terms of growth of patent-to-R&D, the EU28 experienced the second sharpest decline between 2005 and 2012 (-38%) after China (-60%) due to a drop in the number of patents accompanied by an increase in R&D expenditure.

— In **food, agriculture, bioeconomy,** the number of patents per R&D expenditure as estimated for 2012 was the highest in the EU28 (2.36) followed by South Korea (1.57), Japan (1.21) and the US (1.08). As in the case of Health, South Korea's positioning as second highest in the patent-to-R&D ratio is explained by its low BERD in relative terms (one fifth of the EU28 budget). In the US, BERD is almost twice as high as the BERD in the EU28 while patents are 30% lower than in the EU28. The level of BERD for Japan and China is closer to the EU28 BERD but both have significantly less patents. In terms of growth, the EU28 has experienced the second sharpest decline (-28%) after China (-50%) due to a drop in the number of patents.

— In **energy's** patent-to-R&D ratio, the EU28 (2.71) and US (2.62) ranked at the top as estimated for 2012 and were followed closely by Japan (2.14) and South Korea (2.09). BERD was higher for the EU28 although no significant differences between the countries (except South Korea) existed. China increased its BERD by more than five times and the EU28 and Japan by 37% and 39% respectively. All countries recorded significant growth of patents ranging from a six fold growth for China (maximum) to 56% for the US (minimum). Japan's patents grew by 88% and the EU28's by 93%.

— In **transport**, the patent-to-R&D ratio was the lowest for all countries compared to the rest of the SGCs. The highest values in 2012 were estimated for EU28 (0.18) followed by Japan and South Korea (0.15 each) and the US (0.13). The reason for this low patent-to-R&D ratio was given through particularly higher BERD values compared to the rest of the SGCs, and the fact that patents were not proportionally high.

- In **climate,** one can observe a different picture. Here, the patent-to-R&D ratios were the highest and BERD values the lowest among the SGCs. In 2012, the EU28 ranked first (9.99) followed by the US and South Korea (7.95 and 7.94 respectively). South Korea was the country with the highest growth (54%) and China the one with the largest losses (-49%). In terms of patents, all countries experienced growth over 2005-2012 with China and South Korea experiencing threefold increases in patenting, the EU28 and Japan ca. 26% each and the US 12%.
- In **security** the differences between the countries were not highly pronounced in terms of the patent-to-R&D ratio. In 2012, the EU28 ranked first (1.49) and was closely followed by South Korea (1.34), Japan (0.94), the US (0.86), and China (0.70). The highest growth in this ratio was estimated for South Korea with 46%. In terms of BERD, China and South Korea experienced fourfold and more than twofold growth respectively while in terms of patents both countries experienced significant threefold growth.

Specialisation indices for the EU28 during the period 2010-2012 were calculated based on the relative proportion of R&D expenditure and transnational patents in each SGC in comparison with other countries worldwide (interpretation is similar to indices for KETs presented in section 3.1.3). In terms of both indices and for most SGCs, the EU28 did not show a significant over- or under-specialisation compared to the "world". The exceptions were a low specialisation in terms of R&D in health (index: -31) and only a slightly high specialisation in transport (index: 21).

Figure 26 and **Figure 27** compare the specialisation profile of the EU28 across SGCs with the US, China, Japan and South Korea in terms of, R&D expenditure and transnational patents. Results show that the EU28 was ranked first in transport (both in terms of patents and R&D) and Climate (in terms of patents) although the EU28 was not highly specialised compared to the world.

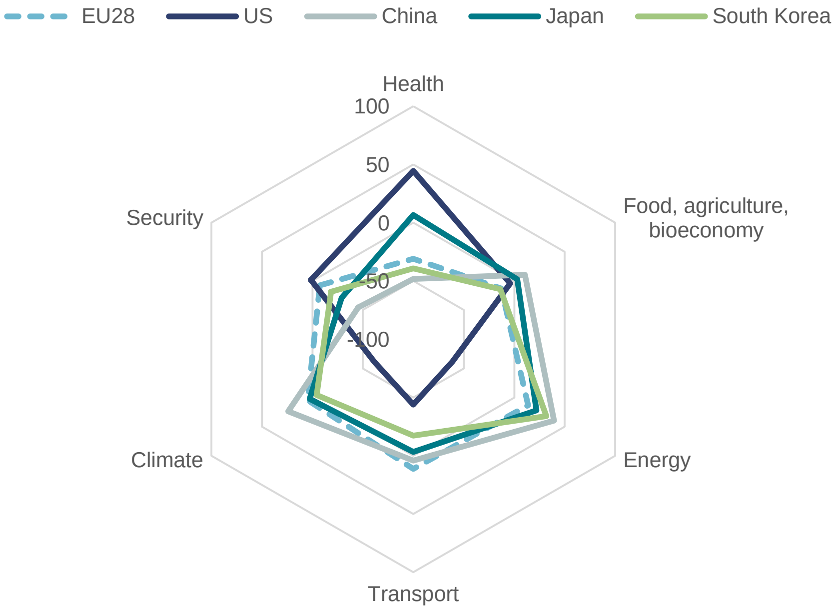
The US showed a strong specialisation especially in health (R&D-based index: 44 and patent-based index: 40). US patents also presented a strong concentration in security (index: 26). Specialisation was low for the US in energy, transport and climate according to shares of R&D and patents in these SGCs.

China showed negative specialisation indexes for patents in all SGCs, which suggests that Chinese patents are particularly underrepresented in SGCs compared to other countries in the world. On the other hand, shares of Chinese R&D in energy and climate are high, with specialisation indices equal to 40 and 24 respectively. In health (index: -48) and security (index: -46) China showed a very low proportion of R&D compared to the "world".

Japan showed a high specialisation of R&D and patents in energy (indices 22 and 24 respectively) and a very low specialisation in security (-29 and -41). Shares of R&D in other SGCs in Japan were similar to other countries worldwide (health index: 7; food, agriculture, bioeconomy index: 3; climate index: 3; and transport index: -3). However, patents in health were underrepresented (index: -44).

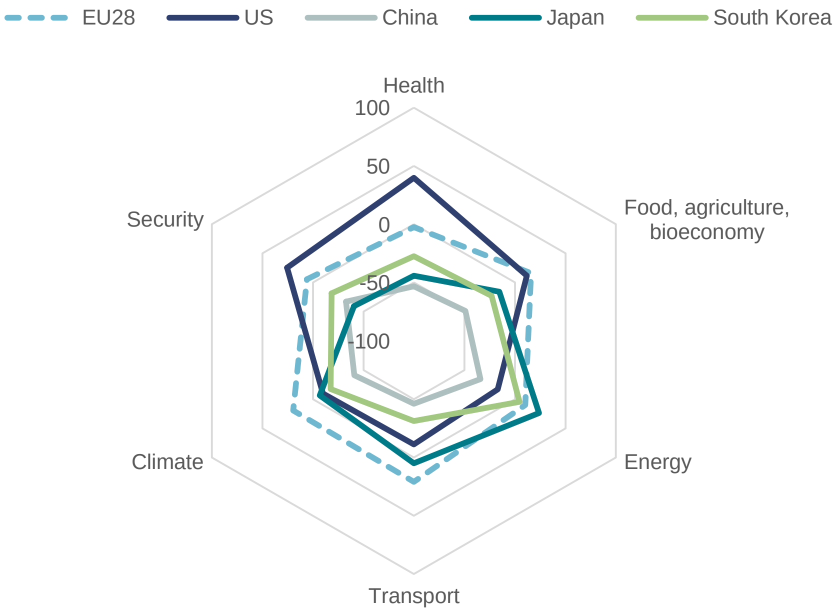
Specialisation indices for South Korea were negative for all SGCs, except for energy, for which specialisation indices were high for R&D (index: 32) and average for patents (index: 5).

Figure 26: Specialisation of EU28 vs. selected countries (based on R&D expenditure)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

Figure 27: Specialisation of EU28 vs. selected countries (based on transnational patents)



Source: OECD – MSTI; EPO – PATSTAT; Fraunhofer ISI calculations.

4 Multivariate models

So far the descriptive statistics provided an overview of the current trends and the positioning of Europe within KETs and SGCs. In this chapter, we will focus on the relationships between R&D, patents and value added in an international comparison. To be more precise, we run a model on the relationship between BERD and patent filings as well as patent filings and value added within KETs and SGCs. This relates to the classical idea of increased input in R&D leading to larger innovative output (as indicated by the number of patent filings), which ultimately leads to an increase in welfare and growth.[18]

Essentially, this relation was already put forward for example in the seminal work by Griliches on patents and R&D (Griliches 1981; 1988; Griliches 1990). This approach has been applied several times since then (see for example Arora et al. 2003; Blind, Frietsch 2003; Cohen et al. 2002; Hingley 1997; Kortum 1997; Reiß et al. 2016). One of the most famous implementations of such an input-output model is the so called CDM-model (Crépon et al. 1998), which uses micro (firm-level) data, however. They link R&D investment to innovation and in a further step innovation to productivity. A recent implementation of this model, expanded to include additional factors, namely management practices, was presented by Bartz et al. (Bartz et al. 2016). An interesting analysis was performed by Ophem et al. (2002) who not only take into account the impact of R&D on patents, but also vice versa, the impact of patents on R&D. They explain it by innovation rents – which generate additional income based on patents – that can be reinvested into further R&D and therefore generate additional knowledge that can be patented afterwards.

This is to derive empirical laws of impact, i.e. the impact of additional R&D expenditures on the amount of patent filings and the impact of patent filings on value added. The results thereby also give an association between R&D and value added. Based on this, we are able to provide more in-depth conclusions and recommendations for political action. A panel data set of the indicators for all countries under analysis here allows for an integrated analysis and perspective on the innovation activities in KETs and SGCs.

4.1 The models

In order to assess the relationship between BERD, patent filings and value added within KETs and SGCs, we ran a series of panel regression models.[19] Based on the data that had already been collected within the project, we first of all created an integrated panel data set[20] for further analysis. It contains data on BERD, patent filings and value added within SGCs and KETs across countries and years. In addition, we included further control variables that are described in more detail below. The panel runs from 2000 to 2014 and includes 27 EU countries[21] and 12 fields (6 KETs, 6 SGCs). In sum, it contains 4,680 observations in 312 country/year/field groups.

In order to answer our questions, a series of models separated by KETs and SGCs will be estimated. We started by taking a closer look at the time delay between BERD and patents as well as patents and value added. This is essential as it provides us with the information on when R&D expenditure is best reflected in subsequent patenting. As soon

[18] We are aware of the fact that this linear innovation model does not mirror reality as innovation processes are much more complex, including feedback loops etc. However, we aim to analyse basic structures and trends of innovation in KETs and SGCs. Therefore, this rather simple model serves as a basic assumption.

[19] We have used patent shares to assign BERD, value added and employment to KETs and SGCs. However, we are still able to run the regression models without running into the problem of a circular logic, as the assignment was done at the micro-level and, in addition, exogenous weights were used for the re-calculation. Therefore, BERD is not direct (linear) transformation of the distribution of patents.

[20] This data set only uses data that have been sent out to the IPTS within the course of the project. To ease the use of the data set, however, variables that were not necessary for the analysis were deleted. In addition, data on country groups as well as the data that are not KETs or SGCs related were dropped.

[21] We only use data for 27 EU countries, excluding Croatia here, where the time series for some indicators (value added and employment) are not as complete as for the others.

as we gathered this information, we knew which time lag needed to be used for our further analyses.

The model is a fixed effects regression model, i.e. a Within Estimator that eliminates the fixed effects by centring each variable on its individual-specific mean, taking into account potentially endogenous individual effects. To decide between fixed or random-effects, we additionally employed a Hausman test, which showed that the random-effects assumption (that explanatory variables are uncorrelated with field-specific effects) is violated. This would lead to systematically biased coefficients as well as standard errors. Therefore, only a fixed-effects estimator results in unbiased estimates (Wooldridge 2002).

The models are specified as follows:

$$Pat_{it} = \alpha_{1it} BERD_{it} + \alpha_{2it-1} BERD_{it-1} + \alpha_{3it-2} BERD_{it-2} + \alpha_{4it-3} BERD_{it-3} + x'_{it}\,\beta_{it} + f_i + u_{it} \qquad (1)$$

$$VA_{it} = \alpha_{1it} Pat_{it} + \alpha_{2it-1} Pat_{it-1} + \alpha_{3it-2} Pat_{it-2} + \alpha_{4it-3} Pat_{it-3} + x'_{it}\,\beta_{it} + f_i + u_{it} \qquad (2)$$

with $i = 1,...,n \qquad t = 1,...T$

where Pat_{it} denotes the number of patent filings of unit i in period t, $BERD_{it}$ and VA_{it} are BERD and value added, x_{it} is a vector of control variables, f_i is a field-specific effect and u_{it} idiosyncratic errors.

The control variables include country-dummies to control for country-specific effects as well as time-dummy variables to control for period-specific effects. We specify the model once with absolute values and once using logarithmic transformations of our main variables, namely BERD, employment, value added and transnational patent applications. The log-transformation has two advantages: First, a simple statistical advantage is that the impact of outliers in the data is restricted. The second advantage, and this is even more important here, is that the assumption of decreasing marginal effects can be modelled by the log-transformation. In other words, the difference between the tenth and the eleventh patent is weighted higher than a difference between the thousandth and the thousand and first patent. Both model implementations – absolute and log-transformed models – will be reported.

In a second series of models, we aim to shed some more light on the interrelation between BERD and patents as well as patents and value added within KETs and SGCs. Here, we also run fixed-effects panel regression models with the following specifications:

$$Pat_{it} = \alpha_{1it} BERD_{it} + \alpha_{2it} RD\text{intens}_{it} + \alpha_{3it} EMP_{it} + \alpha_{4it} COPAT_{it} + x'_{it}\,\beta_{it} + f_i + u_{it} \qquad (3)$$

$$VA_{it} = \alpha_{1it-1} Pat_{it-1} + \alpha_{2it} RD\text{intens}_{it} + \alpha_{3it} EMP_{it} + \alpha_{4it} COPAT_{it} + x'_{it}\,\beta_{it} + f_i + u_{it} \qquad (4)$$

with $i = 1,...,n \qquad t = 1,...T$

where EMP_{it} denotes the number employees of unit i in period t, $COPAT_{it}$, x_{it} again is a vector of control variables, f_i is a field-specific effect and u_{it} idiosyncratic errors. The remaining variables are the same as in (1) and (2). Note that in (4), the one-year lagged version of patent filings is used as the explanatory variable, which is an outcome of the models on the time lags, for which the results can be found below.

Besides controlling for country- and period-specific effects, in these models we further control for size effects by including the number of employees in each country/year/field combination. In addition, we control for the R&D intensity, i.e. R&D per patent, to control for the fact that R&D might be more expensive in one field or another which could influence the outcome in terms of patents and value added. Finally, the number of co-patents is controlled for as cooperation patterns could also have an influence on patenting and value added. Here, we also specify the model once with absolute values and once with log-transformed variables on BERD, patents, employees and value added.

4.2 Summary Statistics

Before we dig deeper into the results of these estimations, we provide summary results of our data in **Table 1**. It shows that the number of observations is largest for patents. The fact that we have only 4,314 values for patents in our data set is due to the fact we do not yet have patent data for the year 2014, so that this country/field combination contains missing values. However, some countries do not file any patents in individual fields in individual years, so that the effective number is lower than the number of possible combinations of 4,536. The mean value of R&D expenditures is 1.5 billion EUR within the KETs and SGCs fields, with a minimum of 160,000 EUR in one country in one field and a maximum of almost 74 billion EUR. The standard deviation is rather large at a level of 6.2 billion. On average in the fields and in the countries analysed, we find 1.1 million employees, generating on average about 70 billion of value added and they do this by filing 820 transnational patents. All variables show rather high standard deviations and large ranges between minimum and maximum. This again justifies the use of the logarithmic transformations. Their average values are also displayed in the summary statistics, but are much harder to interpret. In the regressions, however, their interpretation is much simpler and more straightforward as the reported coefficients will reflect elasticities.

The total BERD-intensity is on average about 1%, which means that the countries (re)invest about 1% of their value added in R&D. The share values are on average at a level of about 7%, meaning that the countries – on average – reach a share of 7% in EU values and while the standard deviation is rather high in terms of BERD, value added and employment, it is much smaller in the case of transnational patents. This proves that patents are much more concentrated than the other factors and – looking at maximum and minimum values – they are also more skewed towards larger countries, mainly in Western and Northern Europe.

Table 1: Summary statistics for EU countries

	Obs	Mean	Std. Dev.	Min	Max
berd	2,400	416	1871	0	32100
emp	3,408	605	1728	0	18637
value_added	3,384	32112	101128	12	1134531
pat_tn	4,314	211	596	0	6928
ln_berd	2,400	3.58	2.33	0.01	10.38
ln_emp	3,408	4.64	2.02	-1.70	9.83
ln_va	3,384	8.34	2.13	2.49	13.94
ln_pat	3,890	3.47	2.17	0.00	8.84
copat_tn	4,206	44	108	0.00	1331
rdintens_total	2,460	0.01	0.01	0.00	0.03
berdshare	2,400	0.06	0.10	0.00	0.60
empshare	3,408	0.07	0.10	0.00	0.47
vashare	3,384	0.07	0.10	0.00	0.46
tnpatshare	4,314	0.07	0.08	0.00	0.67
tncopatshare	4,206	0.08	0.11	0.00	1.00
ctry	4,680	20	12	1	40
field	4,680	7	3	1	12
year	4,680	2007	4	2000	2014

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

4.3 The lag structure

As described above, the effects of one factor on another might not occur immediately, but with a time difference. For the estimation of the models it is important to know these time lags. For example, the effect of business R&D expenditures on patents might be rather straightforward. However, R&D expenditures are spent on R&D projects and the remuneration of researchers within these projects. Some of these projects have rather short time perspectives of a few months, while others might last for several years. As a result, a patent application might emerge, which will be filed. But these patents need to be drafted and filed, before they finally appear in the databases. The impact of R&D (and patents) on output measures like employment or value added is more indirect and results in additional processes that we cannot control for with our data and cannot necessarily fix for all possible cases. While the processes for filing a patent or results are predictable and mostly simple and well-defined procedures, the transfer of R&D into value added is less predictable and the crucial processes are less easy to determine and describe. However, it can still be assumed that there is a time lag in the effect of R&D and patenting on the outcome in terms of value added.

To be able to take these lagged effects into account – that is to say to at least know the size of the "black box" of the transfer effect, given that we cannot look inside – we check and control for the time lags. We do this by running panel regressions with lagged predictors between the different variables/factors we have in our data set.

The results of the lag analysis of the effects of BERD on transnational patent applications are displayed in **Table 2**. Only the two-year and three-year lags are statistically significant in the case of KETs and contemporaneous, one and two-year lags in case of SGCs, but in both cases the significant coefficients point in completely different directions, one being positive and the other being negative to almost the same extent. When we use the logarithmic transformation the one-year lag is statistically significant in case of KETs and also shows the largest coefficient, which suggests using a one-year lag between R&D expenditure and patent applications. In case of SGCs we do not find any significant lag with regard to the relation between BERD and patents.

A one-year lag is most often reported in the empirical literature. For example, based on CIS data, Ophem and Brouwer (2002) find a one-year lag between R&D and patents in time series data. However, they also report a time lag of 3-4 years until patents have an effect on R&D – thus the effect working the other way around. They explain it with the fact that patents lead to successful products and thereby to innovation rents that can be reinvested into R&D. The argument is reasonable, especially as the majority of R&D processes are financed by cash flows and internal funds (Schubert, Rammer 2016).

Frietsch et al. (2014) find no lag between R&D expenditures by companies and their patent applications. They come up with two arguments why this is the case and why it is different to several other empirical findings, e.g. by Ophem and Brouwer. The first argument is a conceptual one and the second a methodological one. First, R&D expenditures by (large) companies are rather stable and they continuously conduct R&D processes. The continuity results from longer term basic or pre-competitive research that does not necessarily lead to large numbers of patents. But there is another kind of R&D that companies conduct, which is more D than R and which more frequently leads to patentable results.

Table 2: Lag analysis of the effect of BERD on patent applications

	Coef.	Std. Err.	t	P>|t|
KETs				
Absolute value: lag of BERD to patents				
berd	0.03	0.02	1.60	0.11
L1.	-0.01	0.02	-0.23	0.82
L2.	0.07	0.02	2.63	0.01
L3.	-0.09	0.02	-4.34	0.00
Obs.	696			
R^2 within	0.07			
F	4.59			
Prob>F	0.000			
Log value: lag of BERD to patents				
ln_berd	0.03	0.15	0.22	0.83
L1.	0.38	0.17	2.19	0.03
L2.	0.22	0.18	1.27	0.20
L3.	0.00	0.17	-0.01	0.99
Obs.	611			
R^2 within	0.05			
F	2.54			
Prob>F	0.007			
SGCs				
Absolute value: lag of BERD to patents				
berd	0.06	0.01	4.46	0.00
L1.	-0.06	0.02	-3.63	0.00
L2.	0.07	0.02	4.30	0.00
L3.	0.02	0.01	1.33	0.18
Obs.	714			
R^2 within	0.13			
F	9.61			
Prob>F	0.000			
Log value: lag of BERD to patents				
ln_berd	-0.16	0.11	-1.44	0.15
L1.	0.08	0.12	0.68	0.50
L2.	-0.01	0.13	-0.05	0.96
L3.	0.04	0.11	0.39	0.70
Obs.	688			
R^2 within	0.03			
F	1.62			
Prob>F	0.108			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

Table 3: Lag analysis of the effect of patent applications on value added

	Coef.	Std.	Err.	t
KETs				
Absolute value: lag of patents to value added				
pat_tn	9.47	2.89	3.28	0.00
L1.	10.50	3.42	3.07	0.00
L2.	-6.49	3.34	-1.94	0.05
L3.	12.66	2.56	4.95	0.00
Obs.	1,470			
R² within	0.39			
F	60.53			
Prob>F	0.000			
Log value: lag of patents to value added				
ln_pat	0.03	0.01	5.31	0.00
L1.	0.02	0.01	3.73	0.00
L2.	0.01	0.01	2.10	0.04
L3.	0.00	0.01	-0.49	0.63
Obs.	1,088			
R² within	0.64			
F	118.86			
Prob>F	0.00			
SGCs				
Absolute value: lag of patents to value added				
pat_tn	-0.13	2.97	-0.04	0.97
L1.	15.85	4.42	3.58	0.00
L2.	-8.26	4.58	-1.80	0.07
L3.	8.24	4.01	2.06	0.04
Obs.	1,542			
R² within	0.45			
F	78.61			
Prob>F	0.000			
Log value: lag of patents to value added				
ln_pat	0.05	0.01	7.25	0.00
L1.	0.02	0.01	3.43	0.00
L2.	0.01	0.01	1.95	0.05
L3.	0.01	0.01	2.23	0.03
Obs.	1,385			
R² within	0.68			
F	178.45			
Prob>F	0.000			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

These kinds of processes are usually more short-term and therefore also result in patents more quickly. When we correlate R&D expenditures and patents, we take into account the "fixed" costs of basic or pre-competitive research as well as development activities with shorter durations, which are more "variable costs". If R&D expenditures change, they change at the level of short term activities. It is these short-term (less than one year) "variable costs" that highly correlate with patents. The second argument is that they use – as we do here – the priority year of the patent filings, while many other researchers often use the application date of patents, which is often one year later – due to the patenting procedure. So when we find a zero lag between R&D expenditures and patents based on the priority year, it is effectively the same as a one-year lag in conjunction with the use of the application year. The argument for using the priority year is that this is the earliest date of a patent filing worldwide, which is closest to the date of invention, which in turn is closest to (the end of) the R&D process.

The fact that we do not have clear and significant results in our empirical analysis and a refrain of arguments from the empirical literature, let us conclude that we should use a time lag of zero years in case of BERD and patents.

Table 3 provides the data for the impact of patents on value added. Here we find statistically significant effects for almost all lags and also for both model implementations – the absolute models and also the logarithmically transformed data. This holds for KETs as well as SGCs. Again, the sign of the coefficients changes and the values of these coefficients are almost balanced in all cases. This again leads us to the conclusion that we cannot finally determine the lag structure based on our data. We therefore assume a one-year time lag between patents and value added for our models. The argument for assuming this is rather simple: A patent is not yet a product that can directly be commercialised, but it takes time to implement it in a product, to produce it and to commercialise it. Only then can it take effect on the value added of companies – or on countries in our case. One could well argue that one year might not be enough in several cases. However, if we increase the lag to two years, the effects from other factors (e.g. additional products that were commercialised fast) might bias the results.

4.4 KETs

We now turn to the second series of models, where we aim to shed some more light on the interrelation between BERD and patents, as well as patents and value added within KETs and SGCs. We present a form of sequential modelling approach, starting with the impact of BERD on patent trends and in a second step (described in the next section) the impact of patents on economic performance – measured by value added data. In contrast to the models in section 4.3, we use the independent variables only in the described time-lag structure and add some further variables.

The results of the first models on the impact of BERD on patent applications are displayed in **Table 4** for the Key Enabling Technologies. In the implementation of the models using absolute values, BERD and also the R&D intensity[22] have no significant coefficient, i.e. the P>|t| values are above 0.05, implying that they are not significant at the 5% level. International co-patents are positively related to the number of patent filings, meaning that countries with higher shares of co-patents also increase their numbers of patent applications – all other factors being equal. This finding is in line with most of the empirical literature that relates knowledge flows in international collaborations to technological upgrades for both sides, but more so for the smaller and technologically inferior country. In addition, international co-patents usually have larger patent families, implying that they address more or larger markets and are therefore more likely to be filed on the transnational level, which we are focusing on with our patent analysis. The number of employees shows a slightly negative and significant

[22] The R&D intensity is added to the models in addition to absolute R&D to acknowledge specialization differences (or orientation of a country towards R&D) and not only scale differences in terms of absolute R&D.

coefficient, implying that smaller firms are relatively more active in KETs patenting than large firms. However, this might also be a result of decreasing marginal returns for the larger firms, or the firms that are filing more patents within KETs in absolute terms.

This calls for a logarithmic implementation of the model, which is presented in the lower panel of **Table 4**. In this specification, BERD as well as the total R&D intensity of the countries show statistically significant coefficients. Co-patents keep their significant impact, while employment – although still showing a negative coefficient – is not significant any more. These coefficients can be interpreted as elasticities, i.e. an increase of (logged) business R&D by 1% increases the number of (logged) patents by 0.81% in the case of KETs in Europe. The simple conclusion of this finding is: if the number of patents is to be increased, the R&D input should/could be increased as one reasonable factor. However, it has to be kept in mind that there are decreasing marginal returns, implying that the yield of additional R&D expenditures tends to get smaller with increasing expenditures.

Table 4: Models of the impact on patent applications, absolute and logarithmicimplementation

KETs				
absolute				
	Coef.	Std. Err.	t	P>|t|
berd	0.01	0.01	0.61	0.54
rdintens_total	22.01	415.94	0.05	0.96
emp	-0.11	0.04	-2.53	0.01
copat_tn	1.55	0.08	19.42	0.00
2006	3.90	2.07	1.89	0.06
2007	6.78	2.08	3.26	0.00
2008	5.79	2.15	2.69	0.01
2009	6.51	2.22	2.94	0.00
2010	6.87	2.17	3.17	0.00
2011	7.68	2.29	3.35	0.00
2012	8.54	2.38	3.58	0.00
_cons	85.74	11.32	7.58	0.00
Obs.	1,122			
R^2 within	0.31			
F	39.56			
Prob>F	0.000			
LOGs				
ln_berd	0.81	0.11	7.66	0.00
rdintens_total	-28.88	12.38	-2.33	0.02
ln_emp	-0.12	0.23	-0.50	0.62
copat_tn	0.01	0.00	5.53	0.00
2006	0.01	0.05	0.22	0.82
2007	0.00	0.06	0.04	0.97
2008	-0.06	0.06	-1.02	0.31
2009	0.08	0.06	1.29	0.20
2010	0.05	0.06	0.78	0.44
2011	0.02	0.06	0.34	0.74
2012	0.00	0.07	-0.03	0.97
_cons	0.43	1.12	0.38	0.70
Obs.	1,001			
R^2 within	0.15			
F	13.60			
Prob>F	0.000			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

Table 5: Models of the impact on value added, absolute and logarithmic implementation

KETs				
absolute				
value_added	Coef.	Std. Err.	t	P>|t|
pat_tn L1	11.04	2.71	4.08	0.00
rdintens_total	-42943	38142	-1.13	0.26
emp	64.50	4.04	15.96	0.00
copat_tn	10.68	8.43	1.27	0.21
2006	559.24	205.93	2.72	0.01
2007	990.86	206.01	4.81	0.00
2008	934.27	212.06	4.41	0.00
2009	392.68	218.67	1.80	0.07
2010	1213.08	211.89	5.72	0.00
2011	1641.25	220.01	7.46	0.00
2012	1939.32	224.28	8.65	0.00
_cons	-4604.73	1108.70	-4.15	0.00
Obs.	1,146			
R² within	0.31			
F	40.79			
Prob>F	0.000			
LOGs				
ln_pat L1	0.02	0.01	3.39	0.00
rdintens_total	-2.54	2.03	-1.26	0.21
ln_emp	0.78	0.05	16.67	0.00
copat_tn	0.00	0.00	-0.32	0.75
2006	0.07	0.01	6.47	0.00
2007	0.13	0.01	12.10	0.00
2008	0.16	0.01	14.88	0.00
2009	0.09	0.01	8.13	0.00
2010	0.16	0.01	14.00	0.00
2011	0.20	0.01	17.54	0.00
2012	0.21	0.01	18.10	0.00
_cons	4.62	0.22	20.52	0.00
Obs.	1,002			
R² within	0.51			
F	79.67			
Prob>F	0.000			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

In **Table 5**, the impact of patents and other factors on the value added in KETs are depicted. The direct impact of BERD is not included here as we have modelled that in the previous regression on the connection between BERD and patents. BERD is thus indirectly covered by the patents variable within this regression. Indeed, we find a positive and significant correlation of patents – lagged by one year – with value added. While the R&D intensity of a country does not seem to have a statistically significant relation to the value added in the KETs fields, employment does. With every 1,000 employees the value added increases by 64 million EUR – on average. International co-patents do not have a significant impact in this model.

Specifying the core factors of the model as log transformed variables does not change the overall results. Lagged patent applications and employment keep their statistically significant positive effects.

Taking the two regression estimations together, we find a general proof of our modelling approach for the KETs fields. At least in the case of the logarithmic transformation we find a positive effect of BERD and of international collaboration on the patenting activities. In a second estimation step, we then find a positive and statistically significant effect of patents and employment on value added. Therefore we can conclude that, based on our models, it is justified to increase the R&D expenditures in KETs fields to increase the value added mediated by patent applications.[23] Yet, it once again has to be stressed that decreasing marginal returns do occur, implying that initial effects of R&D on innovation and ultimately value added diminish as the additional input rises.

4.5 SGCs

We have seen in the chapter on the descriptive statistics that SGCs cover a much larger share of the European economy than KETs. All indicators for SGCs show rather high shares in total economic indicators. In addition, KETs are seen as a relevant and considerable input to performance in SGCs. We have seen in the previous section that KETs, mediated by patents, are related to the economic output in Europe. In this section, we estimate regression models on the relation of SGCs to the economic output, similar to the models presented above.

Table 6 provides the results of the regression models where we regressed BERD and other control factors on patent applications in SGCs. It can be observed from the table that business R&D expenditures and international collaboration and knowledge flows, as indicated by the international co-patents, are significantly positive related to innovative output as measured by transnational patent filings.

[23] It might thereby also result in growth and jobs in general, as it is intended in the Innovation Union Strategy. However, we did not directly analyse the impact on jobs. It could well be that the technological progress is of a labour-saving nature. Then the direct effect might be a decrease of the number of jobs. In a medium term, however, an indirect effect could be jobs growth as increasing output will trigger more demand, also in other sectors.

Table 6: Models of the impact on patent applications, absolute and logarithmic implementation

SGCs				
absolute				
pat_tn	Coef.	Std. Err.	t	P>|t|
berd	0.02	0.01	3.19	0.00
rdintens_total	-823.14	1299.23	-0.63	0.53
emp	-0.01	0.03	-0.23	0.82
copat_tn	2.24	0.06	35.13	0.00
2006	5.59	6.35	0.88	0.38
2007	9.50	6.47	1.47	0.14
2008	13.21	6.67	1.98	0.05
2009	13.04	6.69	1.95	0.05
2010	23.73	6.61	3.59	0.00
2011	19.43	6.94	2.80	0.01
2012	21.61	7.05	3.06	0.00
_cons	187.39	28.87	6.49	0.00
Obs.	1,182			
R^2 within	0.57			
F	123.91			
Prob>F	0.000			
LOGs				
ln_berd	0.12	0.07	1.63	0.10
rdintens_total	-15.14	9.39	-1.61	0.11
ln_emp	-0.37	0.23	-1.63	0.10
copat_tn	0.00	0.00	4.90	0.00
2006	0.06	0.04	1.48	0.14
2007	0.15	0.04	3.43	0.00
2008	0.22	0.05	4.74	0.00
2009	0.17	0.04	3.87	0.00
2010	0.15	0.05	3.33	0.00
2011	0.25	0.05	5.25	0.00
2012	0.24	0.05	4.88	0.00
_cons	5.47	1.18	4.64	0.00
Obs.	1,135			
R^2 within	0.10			
F	9.96			
Prob>F	0.000			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

When implementing the model after the logarithmic transformation of the core factors, BERD loses some of its predictive power and is slightly below a significance level of 10%. Employment also is close to significant at the 10% level, but with a negative coefficient, which might in part reflect the decreasing marginal size effects that are mainly controlled by this variable. Larger countries and larger SGCs tend to have less additional patent filings, after controlling for the other factors in our model. However, this might also be related to the fact that there is multicollinearity in our data, i.e. the predictor variables are not only correlated to the dependent variable but also with each other, which increases the standard errors. This is not a problem for the model as such, but effects of the variables can partially cancel each other out, leading to non-significant results.

The models of the impact of patents and other control factors on the economic performance in Societal Grand Challenges, reflected by value added, are depicted in **Table 7**. The one-year lagged patent applications have a positive and statistically significant impact on the value added. Each patent increases the value added in SGCs on average by about 19.8 million EUR per year. The logarithmic implementation suggests that a 1% increase in patents results in a 0.03% increase of value added (covered in million EUR). Employment also has a positive effect: every additional 1,000 employees in a SGC are accompanied by an increase of value added by about 67.6 million EUR. The overall business R&D intensity of a country – calculated as the share of business R&D over value added in total industry – has a significant effect only in the logarithmic implementation, but shows negative signs in both models.

Table 7: Models of the impact on value added, absolute and logarithmic implementation

SGCs				
absolute				
value_added	Coef.	Std. Err.	t	P>|t|
pat_tn L1	19.84	3.05	6.51	0.00
rdintens_total	-106383	170574	-0.62	0.53
emp	67.59	3.46	19.51	0.00
copat_tn	-30.93	9.87	-3.13	0.00
2006	1398.60	903.64	1.55	0.12
2007	2851.70	921.83	3.09	0.00
2008	1779.80	959.72	1.85	0.06
2009	447.85	952.51	0.47	0.64
2010	3067.60	939.51	3.27	0.00
2011	4521.70	975.99	4.63	0.00
2012	5399.56	985.31	5.48	0.00
_cons	-20914	3888	-5.38	0.00
Obs.	1,206			
R^2 within	0.36			
F	54.18			
Prob>F	0.000			
LOGs				
ln_pat L1	0.03	0.01	4.11	0.00
rdintens_total	-5.75	1.70	-3.39	0.00
ln_emp	0.62	0.05	12.72	0.00
copat_tn	0.00	0.00	0.16	0.87
2006	0.08	0.01	8.23	0.00
2007	0.15	0.01	16.09	0.00
2008	0.20	0.01	19.97	0.00
2009	0.16	0.01	15.98	0.00
2010	0.20	0.01	21.33	0.00
2011	0.24	0.01	24.66	0.00
2012	0.26	0.01	25.27	0.00
_cons	5.53	0.26	21.60	0.00
Obs.	1,163			
R^2 within	0.62			
F	146.50			
Prob>F	0.000			

Source: EPO – PATSTAT; OECD – STAN; Eurostat; calculations by Fraunhofer ISI.

The reason might be that, given we are talking about technology-based fields, those countries with a high R&D intensity operate in a different value creation model than the ones with low shares. Traditional Schumpeter-markets are addressed with KETs and also with SGCs, which require R&D investments. These investments, however, decrease the value added in the sectors compared to the non-R&D-intensive products (Heckscher-Ohlin markets). In addition, those with low shares might focus and specialise on one or a few SGCs, while larger and generally innovation-oriented countries tend to have broader technology portfolios.

4.6 Summarising conclusions from the regression models

Summing up the findings from the models, we can state that BERD is positively related to patents and patents have a positive effect on the economic output measured by value added in the fields of Key Enabling Technologies. In the case of SGCs, the relationship is more ambiguous and we do not find these direct connections in the same way, as BERD does not explain the patent filings in a statistically significant manner. This implies that in the more technology-based areas of SGCs, R&D and patents are still the most important and market-securing factors in the innovation- and especially in the economic processes. When existing European strengths are concerned, R&D and patents are still necessary and (mostly) sufficient. When it comes to newly emerging, more complex, and enabling fields of technology like in the case of KETs, R&D and patents are still a necessary, but no longer a sufficient means to the end of economic performance.

5 References

Arora, A.; Ceccagnoli, M.; Cohen, W.M. (2003): R&D and the patent Premium, NBER Working Paper Series.

Bartz, W.; Mohnen, P.; Schweiger, H. (2016): The role of innovation and management practices in determining firm productivity in developing economies, UNU-MERIT Working Papers No- 2016-034, Maastricht: UNU-MERIT.

Blind, K.; Frietsch, R. (2003): Sector-based Multidimenisional Forecasting Model (SEFORM), Report to the European Patent Office within the Research Programme: Improvement of Methods for Forecasting Patent Filings: Fraunhofer ISI.

Cohen, W.M.; Goto, A.; Nagata, A.; Nelson, R.R.; Walsh, J.P. (2002): R&D spillovers, patents and the incentives to innovate in Japan and the United States. In: Research Policy, 31, pp. 1349-1367.

Crépon, B.; Duguet, E.; Mairesse, J. (1998): Research, innovation and productivity: an econometric analysis at the firm level. In: Economics of Innovation and New Technology, 7, pp. 115-158.

Dernis, H.; Dosso, M.; Hervas, F.; Millot, V.; Squicciarini, M.; Verzzani, A. (2015): World Corporate Top R&D Investors: Innovation and IP bundles. A JRC and OECD common report , Luxembourg: Publications Office of the European Union.

DTI; IDEA Consult; IFo Institute; ECORYS (2013): Study on the international market distortion in the area of KETs: A case analysis, Rotterdam: ECSIP Consortium.

Du Plessis, M.; Van Looy, B.; Song, X.; Magerman, T. (2009): Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation, EUROSTAT Working Paper and Studies Luxembourg.

European Commission (2013): Innovation Union. A pocket guide on a Europe 2020 initiative, Luxembourg: Publications Office of the European Union.

Frietsch, R.; Neuhäusler, P.; Michels, C.; Bethke, N.; Rothengatter, O.; Eckl, V.; Kladroba, A.; Stenke, G. (2014): Identifikation der Technologieprofile von FuE-betreibenden Unternehmen anhand eines Matchings von FuE- und Patentdaten, Karlsruhe/Essen: Fraunhofer ISI / Stifterverband.

Frietsch, R.; Neuhäusler, P.; Rothengatter, O. (2013): Which Road to Take? Filing Routes to the European Patent Office. In: World Patent Information, 35 (1), pp. 8-19.

Frietsch, R.; Neuhäusler, P.; Rothengatter, O.; Jonkers, K. (2016): Societal Grand Challenges from a technological perspective – Methods and identification of classes of the International Patent Classification IPC, Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis No. 53, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI.

Frietsch, R.; Schmoch, U. (2010): Transnational Patents and International Markets. In: Scientometrics, 82, pp. 185-200.

Griliches, Z. (1981): Market Value, R&D and Patents. In: Economics Letters, 12 (3), pp. 183-187.

Griliches, Z. (1988): Technology, Education, and Productivity. Early papers with Notes to Subsequent Literature, Oxford.

Griliches, Z. (1990): Patent Statistics as Economic Indicators: A Survey. In: Journal of economic literature, 28, pp. 1661-1707.

Grupp, H. (1998): Foundations of the Economics of Innovation - Theory, Measurement and Practice, Cheltenham: Edward Elgar.

Grupp, H.; Münt, G.; Schmoch, U. (1996): Assessing Different Types of Patent Data for Describing High-Technology Export Performance. In: OECD (ed.): Innovation, Patents and Technological Strategies: Proceedings of a Workshop held at OECD Headquarters in Paris, 8-9 December 1994. Paris: OECD, pp. 271-287.

Hernández, H.; Tübke, A.; Hervás, F.; Vezzani, A.; Amoroso, S.; Dosso, M. (2014): The 2013 EU Industrial R&D Investment Scoreboard, European Commission No. Report EUR 26221 EN.

Hingley, P. (1997): The Extent to which National Research and Development Expenditures Affect First Patent Filings in Contracting States of the European Patent Convention. In: World Patent Information, 19 (1), pp. 15-25.

IDEA Consult; ZEW; NIW; TNO; CEA; Ecorys Fraunhofer ISI (2015): Key Enabling Technologies - First annual report: European Commission, DG Growth.

IDEA Consult; ZEW; TNO; CEA (2012): Feasibility study for an EU Monitoring Mechanism on Key Enabling Technologies: Report for the European Commission, DG Enterprise and Industry.

Japan Patent Office (JPO) (2017): The Five IP Offices (IP5) (JPO /EPO /KIPO /SIPO /USPTO). Online: https://www.jpo.go.jp/torikumi_e/kokusai_e/five_ip_offices.htm.

Kortum, S. (1997): Research, Patenting, and Technological Change. In: Econometrica, 65, pp. 1389-1419.

Larsen, P.B.; Van de Velde, E.; Durinck, E.; Piester, H.N.; Jakobsen, L.; Shaprio, H. (2011): Cross-sectoral Analysis of the Impact of International industrial Policy on Key Enabling Technologies, Brussels: European Commission, DG Enterprise and Industry.

Magerman, T.; Grouwels, J.; Song, X.; Van Looy, B. (2009): Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization, EUROSTAT Working Paper and Studies Luxembourg.

Neuhäusler, P.; Frietsch, R.; Mund, C.; Eckl, V. (2016): Identifying the technology profiles of R&D performing firms - A matching of R&D and patent data. In: International Journal of Innovation and Technology Management, DOI: http://dx.doi.org/10.1142/S021987701740003X.

Neuhäusler, P.; Frietsch, R.; Bethke, N.; Mund, C. (2015): Assessing companies' capability to develop advanced manufacturing technologies in selected industrial sectors, Report to IPTS in the AMCAP project, Karlsruhe: Fraunhofer ISI.

Ophem, H.v.; Brouwer, E.; Kleinknecht, A.; Mohnen, P. (2002): The Mutual Relation between Patents and R&D. In: Kleinknecht, A.; Mohnen, P. (eds.): Innovation and Firm Performance. New York: Palgrave, pp. 56-72.

Peeters, B.; Song, X.; Callaert, J.; Grouwels, J.; Van Looy, B. (2009): Harmonizing harmonized patentee names: an exploratory assessment of top patentees, EUROSTAT Working Paper and Studies Luxembourg.

Reiß, T.; Frietsch, R.; Schubert, T.; Kukk, P.; Van de Velde, E. (2016): Study on EU Positioning: An Analysis of the International Positioning of the EU Using Revealed Comparative Advantages and the Control of Key Technologies, EUROPEAN COMMISSION, Directorate-General for Research & Innovation, Brussels: European Union.

Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Office (WIPO), Karlsruhe: Fraunhofer ISI.

Schmoch, U.; Gauch, S. (2004): Innovationsstandort Schweiz, Studie für das Bundesamt für Berufsbildung und Technologie, Karlsruhe: Fraunhofer-ISI.

Schubert, T.; Rammer, C. (2016): Concentration on the Few? R&D and Innovation in German Firms 2001 to 2013, Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis Karlsruhe: Fraunhofer ISI.

Wooldridge, J.M. (2002): Econometric analysis of cross section and panel data, Cambridge, Mass.: MIT Press.

List of figures

List of tables

List of Annexes

Annexes

Annex I - Additional Plots

Figure A1: KETs – BERD

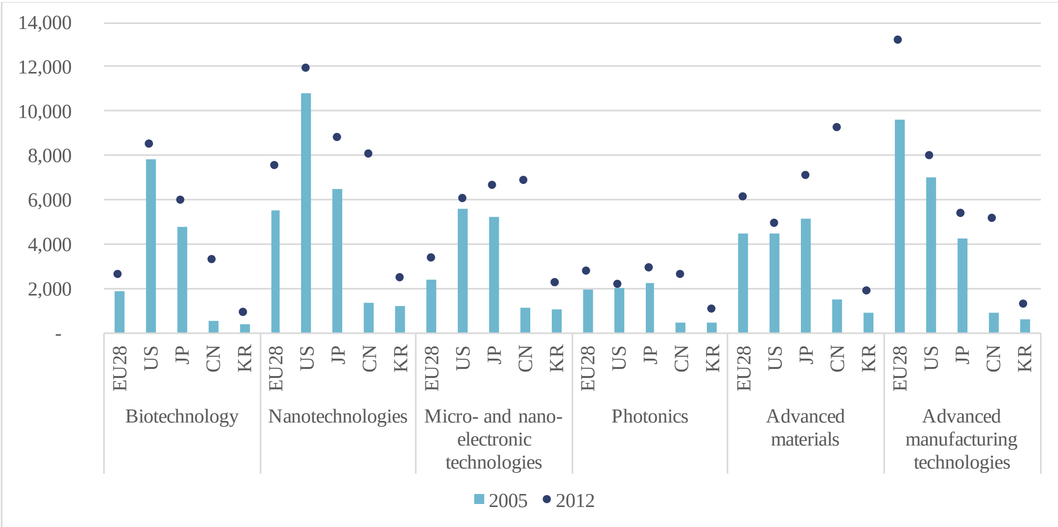


Figure A2: KETs – Patents

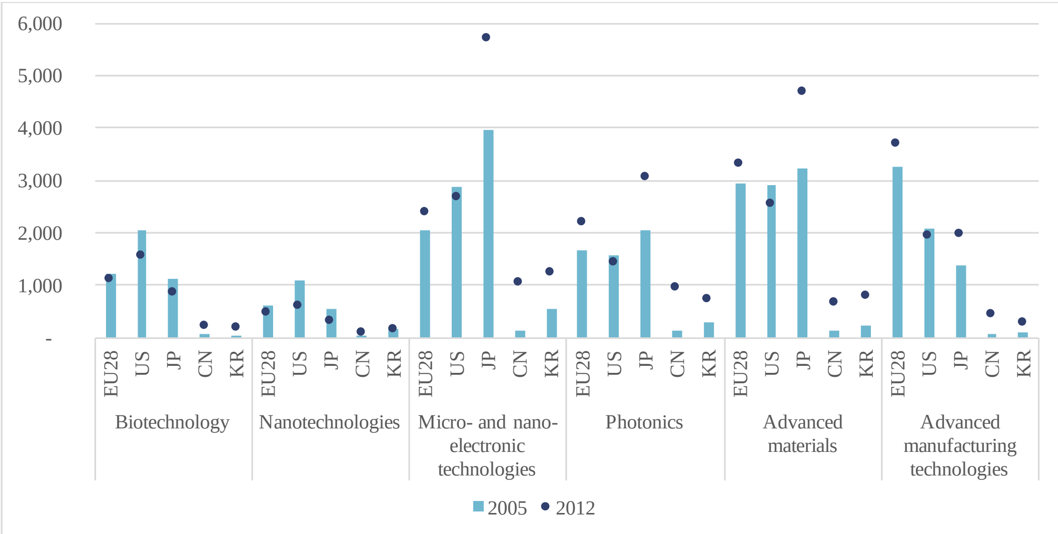


Figure A3: SGC – BERD

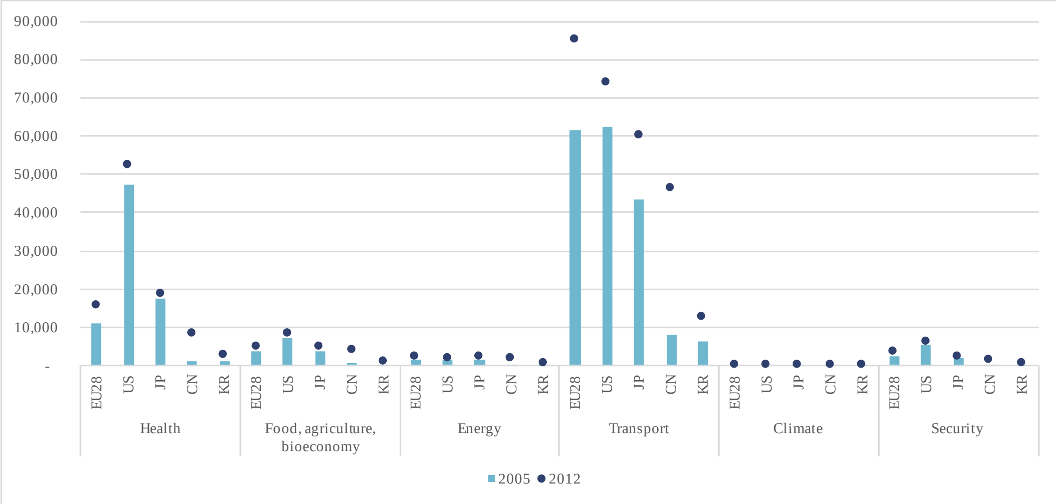
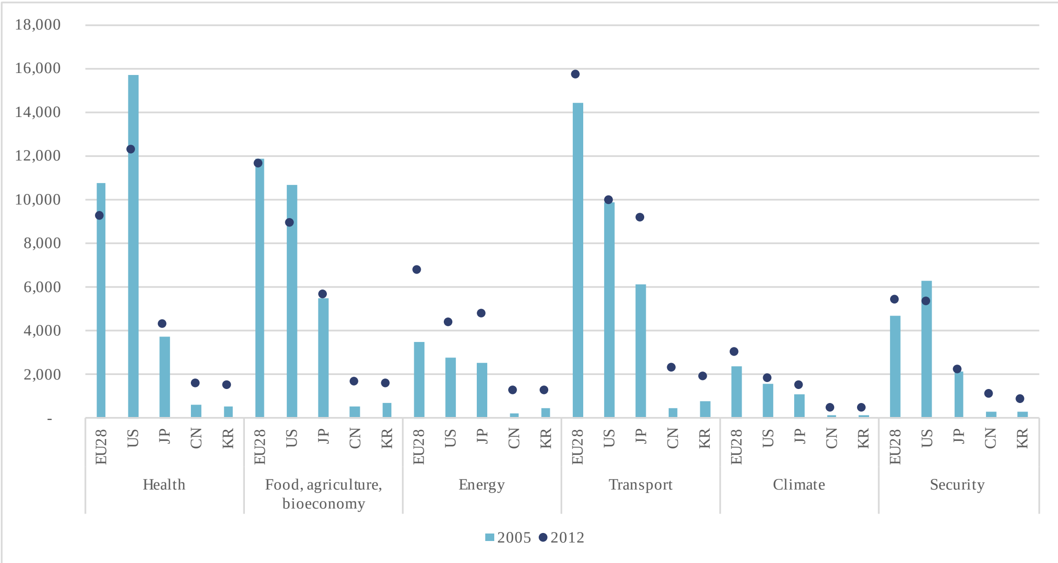


Figure A4: SGC – Patents



Annex II- Methodology Report

Introduction

The study aims to collect and analyse the input and output of business R&D activities in specific sectors, fields, and areas. Mainly R&D investment and patent data are used, supplemented by additional information on the sectoral level, for example on business demography, value added, or employment. One of the core tasks of this study is to collect data and information beyond that which exists in publicly available databases at OECD or Eurostat. This is done, on the one hand, by using alternative sources where available, but, on the other hand, also by systematically and methodically imputing data for certain countries, sectors and/or years. Another core task is to provide patent data by technology as well as sector. For this purpose, technology definitions need to be used (or newly elaborated) and a method is required to convert patent data into sectors. Finally, a method to classify R&D (and other sectoral data) by technologies is also a prerequisite to fulfilling all the tasks.

This is the first inception report on the methodological approaches applied in the project. This document mainly describes the data collection methods for economic and BERD data, including the imputation methods and the strategies to gather additional data (not directly available at EUROSTAT or OECD). It further describes the selection of technology definitions for the patent analyses. This section contains a synthesis of the relevant literature in this context and describes the approach taken to redefine the Societal Grand Challenges. Another section discusses the methodological approach to provide patent data by a sectoral classification (including a synthesis of the relevant literature) and – as a major challenge – to re-calculate R&D data by technological fields. One chapter deals with the data format and the suitability of the data provision to be used as input into the RIO as well as the issue of data updates during the project. The report also describes the methods to produce sector ID cards.

R&D data and other data collection at sectoral level

The first task is to collect **research and development (R&D)** data by industry (BERD) for a number of countries from 2005 onwards and differentiated by industrial sector (NACE). While the so-called Frascati manual published by the OECD defines R&D and presents guidelines on how to properly and thoroughly survey the data, not all countries follow this manual or collect or publish R&D data according to this standard. All EU member states, the associated countries (including EFTA) and OECD member countries provide statistical data in line with the Frascati manual guidelines. However, some countries do not annually report their R&D data or do not publish the full range of industrial sector R&D data. Therefore, there are missing values even in the statistics of the European and OECD countries. Beyond these countries, there are a number of countries, among them Brazil, China and India, which do not provide R&D data in this fashion. And even if they follow the Frascati guidelines, they still might not publish their R&D data by industrial sector. The basic idea underlying this task is to use alternative (i.e. national) sources to fill these gaps. Where such data is not available, gaps have to be filled using imputation methods. One issue is the comparability of data: as data have been collected following specific rules, it is not appropriate to fill in gaps using sources which may have used different procedures, i.e. do not follow the Fascati guidelines.

Apart from BERD data, **other data** need to be collected. Indicators to be calculated/ provided as part of this task and stored in the database are:

- Total BERD,
- BERD and number of researchers (FTE) by sectors
- Share of high-growth enterprises by sectors,
- Share of young high-growth enterprises by sectors,
- Birth, death, and survival rates of business enterprises by sectors,
- Number and share of employees with tertiary education by sectors.

For ratio indicators (shares, rates), numerators and denominators will also be collected separately.

Separate to the database, the R&D investments, the R&D intensity (as % of sales) and the respective growth rates will be provided for the 50 main companies engaged in KETs worldwide.

Sectors and groups of sectors to be covered (from the ToR) are shown in **Table A1**:

Table A1: Sectors

Nace 2.0	**Description**
01-03	Agriculture, forestry and fishing
05-09	Mining and quarrying
10-12	Manufacture of food products; beverages and tobacco products
13-15	Manufacture of textiles, wearing apparel, leather and related products
16-18	Manufacture of wood and paper products; printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22-23	Manufacture of rubber and plastic products and other non-metallic mineral products
24-25	Manufacture of basic metals and fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment n.e.c.
29-30	Manufacture of transport equipment
31-33	Other manufacturing; repair and installation of machinery and equipment
35	Electricity, gas, steam and air conditioning supply
36-39	Water supply, Sewerage, waste management, remediation activities
41-43	Construction
45-47	Wholesale and retail trade; repair of motor vehicles and motorcycles
49-53	Transportation and storage
55-56	Accommodation and food service activities
58-60	Publishing, motion picture, video, television programme production; sound recording, programming and broadcasting activities

Nace 2.0	Description
61	Telecommunications
62-63	IT and other information services
64-66	Financial and insurance activities
68	Real estate activities
69-71	Professional, scientific and technical activities
72	Scientific research and development
77-82	Administrative and support service activities
84	Public administration and defence; compulsory social security
85	Education
86	Human health activities
87-88	Residential care activities and social work activities without accommodation
90-93	Arts, entertainment and recreation
94-98	Other service activities; activities of households as employers and extraterritorial organisations and bodies
99	Activities of extraterritorial organizations and bodies

Source: TOR

All indicators will not be available for all sectors. In particular, structural business indicators are not available for primary sectors and a few tertiary sectors. This issue is addressed in section "Data from Eurostat and OECD".

Data will also be aggregated at the following meta-sector level:

Table A2: Meta-sectors

Meta-sector	NACE Rev. 2 – 2-digit level
High-tech manufacturing	21, 26
Medium high-tech manufacturing	20, 27, 28, 29, 30
Medium low-tech manufacturing	19, 22, 23, 24, 25
Low-tech manufacturing	10-18, 31, 32
Knowledge-intensive services (KIS)	50, 51, 58-66, 69-75, 78, 80, 84-93
Less knowledge-intensive services (LKIS)	45, 46, 47, 49, 52, 53, 55, 56, 68, 77, 79, 81, 82, 94-99
Other	01-19, 33-43

Source: TOR; Eurostat indicators on high-tech industry and knowledge-intensive services, Annex 3 – high-tech aggregation by NACE Rev. 2

If data for meta-sectors are not directly available, NACE sectors will be aggregated into meta-sectors according to the Eurostat correspondence tables.[24] Concerning the manufacturing sectors (high-tech, medium high-tech, medium low-tech and low-tech), the Eurostat aggregation table offers two approaches one using only NACE 2-digit and one also using some NACE 3-digit sectors. Concerning R&D, data are only available at the 2-digit level (see **Table A2**). **In consequence, we will only use the aggregation tables based on NACE 2-digit sectors for all indicators.**

[24] http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf

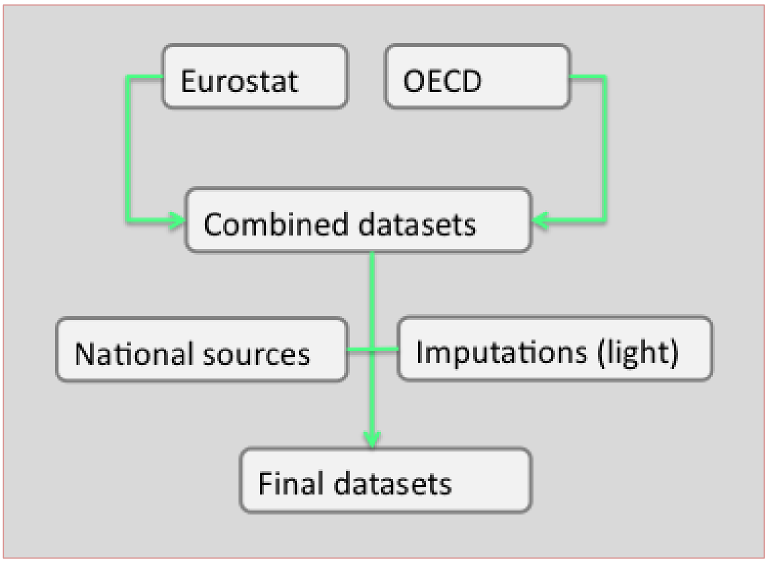
Countries for which data will be collected are EU28 countries, EFTA countries (Iceland, Norway, Switzerland), United States, China, Japan, South Korea, Brazil, India, Israel and Canada.

Quantitative data collection at sector level

Data collection process

Sectoral data requirements as specified in the TOR include indicators readily available from Eurostat and the OECD. Hence, Eurostat and OECD are the main sources used to produce the final datasets. National sources will be used as alternative sources to fill in gaps by scanning the websites of the national statistical offices and directly contacting these offices. Any remaining missing values will be imputed using different imputation methods.

Figure A5: Data collection process



Source: Technopolis

Data from Eurostat and OECD

The EU and OECD databases cover the EU-28 countries, plus, in regard to some indicators, EFTA countries (Switzerland, Norway, Iceland), the BRIC countries Brazil, Russia, India and China, as well as several competing economies, viz. Japan, the United States, Korea, Israel and Canada. The BRIC countries Brazil and India are not included in Eurostat's sources and are only partially available in the OECD databases. Eurostat is the main source of data for EU countries, while OECD is used for countries not covered by Eurostat. It is less likely that data from these two sources will be combined for a given country, but we will examine this possibility.

R&D data on EU member states published by the EU and the OECD are equally incomplete, with gaps of different size and quality. The main reasons for this are:

1. Insufficient data supplied by the member states for specific reasons.
2. Confidentiality: published data should not allow inferences to be made to individual sampling units (i.e. companies). If data are derived from too few companies, countries may refrain from reporting said data. These data may be included in total sums or subtotals, but do not appear as individual data records.

Table A3 lists the sources used to collect sectoral data.

Table A3: Sectoral indicators

Indicator	Description	Main sources
a) Main economic indicators		
Value added	Valued added, gross	Eurostat: nama_10_a64 (B1G) OECD: STAN Database for Structural Analysis
Employment	Total employment	**1** Eurostat: nama_10_a64_e (EMP_DC) OECD: STAN Database for Structural Analysis
b) Business demography		
Number of start-up firms	Number of births of enterprises in t	**2** Eurostat: bd_9n_r2 & bd_9n (V11920) OECD: SDBS Business Demography Indicators
Share of high-growth enterprises	Share of high growth enterprises in the population of active enterprises, measured in employment & turnover number of high growth enterprises measured in employment Number of high growth enterprises measured in turnover	Eurostat: bd_9n_r2 & bd_9n (V97450, V97451, V11950, V11951), bd_9pm_r2 (V11960), bd_9bd_sz_cl_r2 (V16911) OECD: SDBS Business Demography Indicators
Share of young high-growth enterprises	Share of young high growth enterprises (gazelles) measured in employment & turnover number of young high growth enterprises (gazelles) measured in employment number of young high growth enterprises (gazelles) measured in employment	**3** Eurostat: bd_9n_r2 & bd_9n (V97452, V97453, V11952, V11953) OECD: SDBS Business Demography Indicators
Birth rates	Number of enterprise births in the reference period (t) divided by the number of enterprises active in t	**4** Eurostat: bd_9n_r2 & bd_9n (V97020, V11910, V11920) OECD: SDBS Business Demography Indicators
Death rates	Number of enterprise deaths in the reference period (t) divided by the number of enterprises active in t	**5** Eurostat: bd_9n_r2 & bd_9n (V97030, V11910, V11930) OECD: SDBS Business Demography Indicators
Survival rates	Number of enterprises in reference period (t) newly born in t-1 having survived to t divided by the number of enterprise births in t-1	**6** Eurostat: bd_9n_r2 & bd_9n (V97041, V11910, V11941) OECD: SDBS Business Demography Indicators
c) R&D intensity and output		
BERD expenditures	Business enterprise R&D expenditure	**7** Eurostat: rd_e_berdindr2 and rd_e_berdindr OECD: STAN R&D exp. by industry
BERD by type of research		Not available by sector
Researchers	Number of Researchers	Eurostat: rd_p_bempocc and rd_p_bempocc2 (RSE) OECD: Business enterprise R-D personnel by industry
Number and share of employees with tertiary education		Not available by sector
R&D intensity		Derived from above indicators

Source: Technopolis

Eurostat and OECD databases cover all the indicators mentioned in the TOR, with the exception of BERD by type of research as well as employees with tertiary education.

Concerning BERD by type of research, we will have to contact the national statistical offices (which typically conduct the R&D surveys) to obtain the data. We are aware that such data are available in the national surveys of Ireland and Germany – note that questionnaires for national BERD surveys vary by country. The second best solution is to estimate the data based on the available aggregated BERD data by type. Concerning the number and share of employees with tertiary education, we will use data from the Eurostat database on high-tech industry and knowledge-intensive services (Eurostat htec database). Sectors available for this indicator are less granular than at the 2-digit level, but meta-sectors (high-tech manufacturing, knowledge-intensive services, etc.) are available.

When combining data from Eurostat and OECD, the comparability of both sources will be ensured as follows:

- The definition of indicators must be identical in both Eurostat and OECD.
- If Eurostat and OECD data are combined for a given country (which is less likely, as mentioned above), we will compare data that are available in both sources. If they are different, Eurostat data will be given priority over OECD data.
- Comparability is ensured if the guidelines/manuals are the same in both sources. This is the case for R&D (Frascati Manual).

Some indicators do not cover specific sectors. For instance, while the list of sectors in the TOR corresponds exactly to sectors for which data on value added and employment are available in Eurostat, this is not the case for other required indicators. These **missing sectors will not be imputed if no data are available from national sources**. Data at the level of NACE 2-digit manufacturing sectors are not available for employees with tertiary education, but are readily available at meta-sector level in Eurostat.

Table A4: Missing sectors

Nace 2.0	BERD cat	Value Added	Employ-ment	Bus. dem.	R&D	Empl. w. tert. edu.
01-03	**Agriculture, forestry and fishing**	✔	✔		✔	✔
05-09	**Mining and quarrying**	✔	✔		✔	✔
10-12	**Manufacture of food products; beverages and tobacco products**	✔	✔	✔	✔	
13-15	**Manufacture of textiles, wearing apparel, leather and related products**	✔	✔	✔	✔	
16-18	**Manufacture of wood and paper products; printing and reproduction of recorded media**	✔	✔	✔	✔	
19	**Manufacture of coke and refined petroleum products**	✔	✔	✔	✔	
20	**Manufacture of chemicals and chemical products**	✔	✔	✔	✔	
21	**Manufacture of basic pharmaceutical products and pharmaceutical preparations**	✔	✔	✔	✔	
22-23	**Manufacture of rubber and plastic products and other non-metallic mineral products**	✔	✔	✔	✔	
24-25	**Manufacture of basic metals and fabricated metal products, except**	✔	✔	✔	✔	

Nace 2.0	BERD cat	Value Added	Employ-ment	Bus. dem.	R&D	Empl. w. tert. edu.
	machinery and equipment					
26	**Manufacture of computer, electr. and optical prod.**	✔	✔	✔	✔	
27	**Manufacture of electrical equipment**	✔	✔	✔	✔	
28	**Manufacture of machinery and equipment n.e.c.**	✔	✔	✔	✔	
29-30	**Manufacture of transport equipment**	✔	✔	✔	✔	
31-33	**Other manufacturing; repair and installation of machinery and equipment**	✔	✔	✔	✔	
35		✔	✔	✔	✔	✔
36-39	**Water supply, Sewerage, waste management, remediation activities**	✔	✔	✔	✔	✔
41-43	**Construction**	✔	✔	✔	✔	
45-47	**Wholesale and retail trade; repair of motor vehicles and motorcycles**	✔	✔	✔	✔	Wholesale etc.
49-53	**Transportation and storage**	✔	✔	✔	✔	Transp.
55-56	**Accommodation and food service activities**	✔	✔	✔	✔	✔
58-60	**Publishing, motion picture etc.; sound recording, programming and broadcasting activities**	✔	✔	✔	✔	
61	**Telecommunications**	✔	✔	✔	✔	
62-63	**IT and other information services**	✔	✔	✔	✔	
64-66	**Financial and insurance activities**	✔	✔	✔	✔	✔
68	**Real estate activities**	✔	✔	✔	✔	✔
69-71	**Professional, scientific and technical activities**	✔	✔	✔	✔	✔
72	**Scientific research and development**	✔	✔	✔	✔	
77-82	**Administrative and support service activities**	✔	✔	✔	✔	✔
84	**Public admin. and defence; comp. social security**	✔	✔		✔	
85	**Education**	✔	✔	✔	✔	✔
86	**Human health activities**	✔	✔	✔	✔	✔
87-88	**Residential care activities etc.**	✔	✔	✔	✔	
90-93	**Arts, entertainment and recreation**	✔	✔	✔	✔	✔
94-98	**Other service activities; activities of households etc.**	✔	✔	✔	✔	✔
99	**Activities of extraterritorial organizations and bodies**	✔	✔			✔

Source: Eurostat.

Note: empty cell indicates that the variable is not available at sector level

Even when indicators are available in Eurostat/OECD, some can be more problematic than others in terms of their coverage (sector, country, years). We need to bear in mind that, in some cases, significant gaps will remain even after adding national sources and that the final coverage by country, sector and indicator will vary. **Table A5** to **Table A6** present a preliminary overview of data availability for all the indicators. As this reflects work in progress, legends vary depending on the indicator examined.

In brief, gaps in the main economic indicators concerning value added and employment statistics are expected to be the easiest to fill. In contrast, significant data gaps have been identified in the case of business demographics and R&D intensity and output. For example, more than 50% of BERD values are missing across all NACE sectors for EU28 plus Norway and Switzerland in the period 2005-2012. Numerous gaps have also been identified in non-EU countries.

Table A5: Overview of data availability by country, year and indicator: value added and employment

Country	Value added 2005	2006	2007	2008	2009	2010	2011	2012	2013	Employment 2005	2006	2007	2008	2009	2010	2011	2012	2013
European Union (28 countries)																		
Austria	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Belgium	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Bulgaria	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3
Croatia	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	1	1
Cyprus	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Czech Republic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Denmark	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Estonia	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Finland	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
France	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
Germany	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Greece	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Hungary	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ireland	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Italy	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2
Latvia	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Lithuania	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Luxembourg	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Malta	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Netherlands	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Poland	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Portugal	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
Romania	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Slovakia	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Slovenia	2	2	2	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1
Spain	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
Sweden	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
United Kingdom	2	2	2	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2
EFTA (without Liechtenstein)																		
Iceland	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3	3
Norway	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2
Switzerland	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Main Competitors																		
United States	2	2	2	2	2	3	3	3	3	2	2	2	2	2	3	3	3	3
China	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Japan	2	2	2	2	2	3	3	3	3	2	2	2	2	2	3	3	3	3
South Korea	2	2	2	2	2	2	3	3	3	2	2	2	2	2	3	3	3	3
Brazil	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
India	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Israel	2	2	2	2	3	3	3	3	3	2	2	2	2	3	3	3	3	3
Canada	2	2	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3	3

Legend:

1	Full data in NACE Rev.2/ISIC 4
2	NACE 1.1/ISIC 3 available or a few missing sectors
3	No data available

Table A6: Overview of data availability by country, year and indicator: number of start-up firms

Country	Number of start up firms 2005	2006	2007	2008	2009	2010	2011	2012
European Union (28 countries)								
Austria	1	1	1	1	1	1	1	1
Belgium	3	1	1	1	1	1	1	1
Bulgaria	3	1	1	1	1	1	1	1
Croatia	4	4	4	4	4	4	4	2
Cyprus	3	3	3	1	1	1	1	1
Czech Republic	1	1	1	1	1	1	1	1
Denmark	2	2	3	4	2	2	2	2
Estonia	1	1	1	1	1	1	1	1
Finland	1	1	1	1	1	1	1	1
France	2	2	2	2	2	2	2	2
Germany	1	1	1	1	1	1	1	1
Greece	4	4	4	4	4	4	4	4
Hungary	1	1	1	1	1	1	1	1
Ireland	4	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1	1
Latvia	1	1	1	1	1	1	1	1
Lithuania	3	3	1	1	1	1	1	1
Luxembourg	1	1	1	1	1	1	1	1
Malta	4	4	4	4	1	1	1	1
Netherlands	1	1	1	1	1	1	1	1
Poland	2	2	2	2	2	2	2	2
Portugal	1	1	1	1	1	1	1	1
Romania	1	1	1	1	1	1	1	1
Slovakia	1	1	1	1	1	1	1	1
Slovenia	1	1	1	1	1	1	1	1
Spain	2	2	2	2	2	2	2	2
Sweden	2	2	2	2	2	2	2	2
United Kingdom	1	1	1	1	1	1	1	1
EFTA (without Liechtenstein)								
Iceland	4	4	4	4	4	4	4	4
Norway	1	1	1	1	1	1	1	1
Switzerland	2	2	2	2	2	2	2	2
Main Competitors								
United States	3	3	3	3	3	3	3	3
China	4	4	4	4	4	4	4	4
Japan	4	4	4	4	4	4	4	4
South Korea	4	4	4	4	4	4	4	4
Brazil	3	3	4	2	2	2	2	2
India	4	4	4	4	4	4	4	4
Israel	4	4	4	4	4	4	2	2
Canada	2	2	2	2	2	2	2	2

Legend:

1	Full data in NACE Rev.2/ISIC 4
2	Missing sectors in NACE 2/ISIC 4
2	NACE 1/ISIC 3 available
4	No data available

Table A7: Overview of data availability by country, year and indicator: high-growth enterprises

Country	Share of high-growth enterprises								Share of young high-growth enterprises							
	2005	2006	2007	2008	2009	2010	2011	2012	2005	2006	2007	2008	2009	2010	2011	2012
European Union (28 countries)																
Austria	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Belgium	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4
Bulgaria	3	3	3	4	4	4	4	1	4	3	3	4	4	4	4	4
Croatia	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Cyprus	4	4	4	4	4	1	4	4	4	4	4	4	4	1	4	4
Czech Republic	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	1
Denmark	3	3	3	4	2	2	2	4	3	3	3	4	2	2	2	4
Estonia	4	3	3	1	1	1	1	1	3	3	3	1	1	1	1	1
Finland	1	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4
France	4	4	4	4	2	4	4	4	4	4	4	4	2	4	4	4
Germany	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Greece	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Hungary	3	3	3	1	1	1	1	1	4	4	4	1	1	1	1	1
Ireland	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Italy	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	1
Latvia	3	3	3	4	4	4	1	1	3	3	3	4	4	4	1	1
Lithuania	3	3	3	4	4	1	1	1	3	3	3	4	4	1	1	1
Luxembourg	3	3	3	1	1	1	1	4	3	3	3	1	1	1	1	4
Malta	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Netherlands	3	3	3	4	4	4	4	1	3	3	3	4	4	4	4	1
Poland	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Portugal	4	3	2	2	2	2	2	2	4	4	2	2	2	2	2	2
Romania	3	3	3	4	1	1	1	1	3	3	3	4	1	1	1	1
Slovakia	4	4	3	4	4	4	4	1	3	4	4	4	4	4	4	1
Slovenia	3	3	3	2	2	2	2	4	3	3	3	1	1	1	1	4
Spain	4	4	4	2	2	2	2	2	3	3	3	2	2	2	2	2
Sweden	4	4	4	1	4	4	4	4	4	3	3	4	4	4	4	4
United Kingdom	4	4	3	4	4	4	4	1	4	4	4	4	4	4	4	4
EFTA (without Liechtenstein)																
Iceland	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Norway	4	1	4	4	4	4	4	4	4	3	4	4	4	4	4	4
Switzerland	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Main Competitors																
United States	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
China	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Japan	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
South Korea	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Brazil	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4
India	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Israel	4	4	4	1	1	1	1	4	4	4	4	4	4	4	4	4
Canada	1	1	1	4	4	4	4	4	4	4	4	4	4	4	4	4

Legend:

1	Full data in NACE Rev.2/ISIC 4
2	Missing sectors in NACE 2/ISIC 4
2	NACE 1/ISIC 3 available
4	No data available

Table A8: Overview of data availability by country, year and indicator: birth, death and survival rates

	Birth rate								Death rate								Survival rate							
Country	**'05**	**'06**	**'07**	**'08**	**'09**	**'10**	**'11**	**'12**	**'05**	**'06**	**'07**	**'08**	**'09**	**'10**	**'11**	**'12**	**'05**	**'06**	**'07**	**'08**	**'09**	**'10**	**'11**	**'12**
European Union (28 countries)																								
Austria	1	1	1	1	1	1	1	5	1	1	1	1	1	1	5	5	2	2	2	2	2	2	2	2
Belgium	5	3	3	2	2	2	2	2	5	3	3	2	2	2	2	5	5	5	5	2	2	2	2	2
Bulgaria	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	5	4	5	4	2	2	2	2	2
Croatia	5	5	5	5	5	5	5	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Cyprus	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	5	5	4	4	5	2	2	2	2
Czech Republic	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	5	3	3	3	2	2	2	2	2
Denmark	3	3	4	5	2	2	2	2	3	4	4	5	2	2	2	2	3	3	5	5	2	2	2	2
Estonia	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	5	3	3	3	2	2	2	2	2
Finland	3	3	3	1	1	1	1	1	3	3	4	2	2	2	2	2	4	4	4	2	2	2	2	2
France	4	4	4	2	2	2	2	2	4	5	4	2	2	2	2	2	4	4	4	2	2	2	2	2
Germany	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	3	4	4	5	1	1	1	1	1
Greece	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Hungary	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	5	3	3	3	1	1	1	1	1
Ireland	5	5	5	2	2	2	2	2	5	5	5	2	2	2	2	5	5	5	5	2	2	2	2	2
Italy	3	3	3	1	1	1	1	1	3	3	3	1	1	1	1	5	3	3	3	2	2	2	2	2
Latvia	4	4	3	2	2	2	2	2	4	4	3	2	2	2	2	5	4	4	4	2	2	2	2	2
Lithuania	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	5	4	4	4	2	2	2	2	2
Luxembourg	3	3	3	2	2	2	2	2	3	3	3	2	2	2	2	5	3	3	3	2	2	2	2	2
Malta	5	5	5	5	5	1	1	1	5	5	5	5	5	1	1	5	5	5	5	5	5	5	2	2
Netherlands	3	3	3	1	1	1	1	1	3	3	4	1	1	1	1	5	3	3	3	1	1	1	1	5
Poland	4	4	4	2	2	2	2	2	4	4	4	2	2	2	2	5	4	4	4	2	2	2	2	2
Portugal	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	2	2	2	2	2	2	2	2
Romania	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	3	4	4	4	2	2	2	2	2
Slovakia	3	3	3	1	1	1	1	1	3	3	4	1	1	1	1	5	3	3	3	1	1	1	1	1
Slovenia	1	1	1	1	1	1	1	1	3	4	3	1	1	1	1	3	2	2	2	2	2	2	2	2
Spain	3	3	3	2	2	2	2	2	3	3	3	2	2	2	2	5	3	3	3	2	2	2	2	2
Sweden	4	3	3	2	2	2	2	2	4	4	5	2	2	2	2	5	4	4	4	2	2	2	2	2
United Kingdom	4	4	4	1	1	1	1	1	4	4	4	1	1	1	1	5	4	5	4	1	1	1	1	1
EFTA (without Liechtenstein)																								
Iceland	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Norway	3	3	5	1	1	1	1	1	3	5	5	1	1	1	1	3	5	3	5	1	1	1	1	1
Switzerland	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	5	5	5	5
Main Competitors																								
United States	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4
China	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Japan	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
South Korea	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Brazil	5	5	5	3	3	3	3	3	5	5	5	3	3	3	3	5	5	5	5	5	3	3	5	5
India	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Israel	5	5	5	5	5	5	3	3	5	5	5	5	5	5	3	5	5	5	5	5	5	5	5	5
Canada	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5

Legend:

1	Full data in NACE Rev.2
2	Missing sectors in NACE 2
3	ISIC 4 available
4	NACE 1/ISIC 3 available
5	No data available

Table A9: Overview of data availability by country, year and indicator: R&D expenditure and personnel

Country	R&D 2005	2006	2007	2008	2009	2010	2011
European Union (28 countries)							
Austria	5	3	3	5	3	5	3
Belgium	2	2	2	3	3	3	3
Bulgaria	4	3	3	3	3	3	3
Croatia	4	4	4	4	1	1	1
Cyprus	2	2	1	1	1	1	1
Czech Republic	1	1	1	1	1	1	1
Denmark	5	5	4	5	3	3	3
Estonia	4	4	3	3	3	3	3
Finland	4	4	4	3	3	3	3
France	5	5	3	3	3	3	3
Germany	4	4	3	3	3	3	3
Greece	4	5	4	5	5	5	3
Hungary	3	3	3	3	3	3	3
Ireland	2	4	4	4	3	3	3
Italy	4	4	3	3	3	3	3
Latvia	4	4	5	3	3	3	3
Lithuania	4	4	4	4	3	3	3
Luxembourg	5	5	5	5	5	5	5
Malta	2	2	4	3	3	1	1
Netherlands	4	4	4	3	3	3	3
Poland	3	3	3	3	3	3	3
Portugal	4	4	3	3	3	3	3
Romania	4	4	4	3	3	3	3
Slovakia	4	3	3	3	3	3	3
Slovenia	2	2	2	1	1	1	1
Spain	2	2	2	1	1	1	1
Sweden	5	4	3	5	3	5	3
United Kingdom	5	5	3	3	3	3	3
EFTA (without Liechtenstein)							
Iceland	4	4	4	4	4	5	1
Norway	4	4	4	3	3	3	3
Switzerland	5	5	5	3	5	5	5
Main Competitors							
United States	4	4	5	5	5	5	5
China (except H	5	5	5	5	5	5	5
Japan	4	4	4	4	5	5	5
South Korea	2	2	2	2	5	5	5
Brazil	5	5	5	5	5	5	5
India	5	5	5	5	5	5	5
Israel	5	5	5	5	5	5	5
Canada	5	5	5	5	5	5	5

Legend:

1	Full data in NACE Rev.2
2	Full data in NACE 1
3	Missing sectors in NACE 2
4	Missing sectors in NACE 1
5	No data available in Eurostat. Use OECD or national sources.

Table A10: Overview of data availability by country, year and indicator: personnel with tertiary education in high-tech manufacturing

Country	2005	2006	2007	2008	2009	2010	2011	2012	2013
European Union (28 countries)									
Austria	3	3	3	1	1	1	1	1	1
Belgium	3	3	3	1	1	1	1	1	1
Bulgaria	3	3	3	1	1	1	1	1	1
Croatia	4	4	3	1	1	1	1	1	1
Cyprus	4	4	4	1	1	1	1	1	1
Czech Republic	3	3	3	1	1	1	1	1	1
Denmark	3	3	3	1	1	1	1	1	1
Estonia	4	4	4	1	1	1	1	1	1
Finland	3	3	3	1	1	1	1	1	1
France	3	3	3	1	1	1	1	1	1
Germany	3	3	3	1	1	1	1	1	1
Greece	3	3	3	1	1	1	1	1	1
Hungary	3	3	3	1	1	1	1	1	1
Ireland	3	3	3	1	1	1	1	1	1
Italy	3	3	3	1	1	1	1	1	1
Latvia	4	4	4	1	1	2	2	2	2
Lithuania	4	4	4	2	2	2	2	2	2
Luxembourg	4	4	4	2	2	2	2	2	2
Malta	5	5	5	2	2	2	2	2	2
Netherlands	3	3	3	1	1	1	1	1	1
Poland	3	3	3	1	1	1	1	1	1
Portugal	4	4	4	1	1	1	1	1	1
Romania	3	3	3	1	1	1	1	1	1
Slovakia	3	3	3	1	1	1	1	1	1
Slovenia	3	3	3	1	1	1	1	1	1
Spain	3	3	3	1	1	1	1	1	1
Sweden	3	3	3	1	1	1	1	1	1
United Kingdom	3	3	3	1	1	1	1	1	1
EFTA (without Liechtenstein)									
Iceland	4	4	4	2	2	2	2	2	2
Norway	4	3	3	1	1	1	1	1	1
Switzerland	3	3	3	1	1	1	1	1	1
Main Competitors									
United States	6	6	6	6	6	6	6	6	6
China	6	6	6	6	6	6	6	6	6
Japan	6	6	6	6	6	6	6	6	6
South Korea	6	6	6	6	6	6	6	6	6
Brazil	6	6	6	6	6	6	6	6	6
India	6	6	6	6	6	6	6	6	6
Israel	6	6	6	6	6	6	6	6	6
Canada	6	6	6	6	6	6	6	6	6

Legend:

1	Full data in NACE Rev.2
2	Almost complete data in NACE 2
3	Full data in NACE 1
4	Almost complete data in NACE 1
5	Many missing values
6	No data available in Eurostat.

Some countries are not covered in general in the Eurostat/OECD databases with regard to the required indicators. For these countries, if no alternative data source is found (i.e. national statistical websites/databases), we will need to contact the national offices in order to retrieve more information about their sectoral data, even scarce data or data with gaps/missing data. **Table A11** lists the countries not covered by Eurostat/OECD with regard to our indicators.

Table A11: Country with no coverage in Eurostat/OECD

Indicator	**No data available by sector**
a) Main economic indicators	
Value added	HR, BR, CN ,IN
Employment	BG, BR, CH, CN, IN
b) Business demography	
# start-up firms	EL, IS, CN, IN, JP, KR US
Share high-growth firms & young high-growth firms	AT, BE, DE, EL, HR, IE, MT, PL, UK, CH, IS, CN, IN, JP, KR, US. Only young firms missing: BE, UK, IS, BR, CA
# birth/death/survival rates	EL, HR (exc. birth rate), CH (exc. survival rate), IS, CH, IN, JP, KR, US
c) R&D intensity and output	
R&D	IS, BR, CA, CN, IN
Personnel with tertiary education in ht-sectors	CH, IL, BR, CA, IN, JP, KR, US

Note: preliminary overview

National sources

Alternative sources of data are located by scanning the websites of the national statistical offices as well as contacting the offices directly. The comparability of Eurostat/OECD data and data from national sources is addressed as follows:

— Comparability must be checked by examining the definition of the indicators and the guidelines used by the national offices. These are then compared with Eurostat/OECD definitions.

— If no data is available in Eurostat/OECD, only definitions and guidelines will be compared, and the nationally available data used.

In any case, the origin of each data point will be documented in the metafile.

We will rely on the following list of contacts at the national statistical offices (see **Table A12**). Specific national organisations other than the main statistical office may be responsible for the R&D indicators of their country (for instance the Stifterverband in Germany, or the Belgian Science Policy Office in Belgium). This is why we use a tailored list for these indicators as a first contact point. This list includes the national delegates for R&D indicators at the Eurostat/OECD meetings. If these contacts cannot provide us with non-R&D indicators, we will ask them to direct us to the relevant contact persons, or we will contact the main statistical office.

Table A12: Contacts at national statistical offices

	Country	R&D	Other statistics	Website of statistical office
	European Union (28 countries)			
AT	Austria	Andreas.Schiefer@statistik.gv.at	Statistik Austria	http://www.statistik.at/web_en/statistics/index.html
BE	Belgium	ziar@belspo.be, Jeoffrey.MALEKMANSOUR@belspo.be	Statistics Belgium	http://statbel.fgov.be/en/statistics/figures/
BG	Bulgaria	VJeleva@NSI.bg	National Statistics institute	http://www.nsi.bg/en
HR	Croatia	skegrom@dzs.hr, duicv@dzs.hr, emira.becic@mzos.hr	Croatian Bureau of Statistics	http://www.dzs.hr/default_e.htm
CY	Cyprus	pprotopapas@cystat.mof.gov.cy	Statistical Service	http://www.mof.gov.cy
CZ	Czech Republic	martin.mana@czso.cz, vaclav.sojka@czso.cz	Czech Statistical Office	https://www.czso.cz/csu/czso/home
DK	Denmark	jbr@dst.dk, LEN@dst.dk	Statistics Denmark	http://www.dst.dk/en
EE	Estonia	tiina.parson@stat.ee	Statistics Estonia	http://www.stat.ee/en
FI	Finland	ari.leppalahti@stat.fi, Mervi.Niemi@stat.fi	Statistics Finland	http://www.stat.fi/index_en.html
FR	France	Geraldine.seroussi@recherche.gouv.fr, xavier.besnard@insee.fr	National Institute of Statistics and Economic Studies	http://www.insee.fr/en/
DE	Germany	Andreas.Kladroba@stifterverband.de	Federal Statistical Office	https://www.destatis.de/EN/Homepage.html
EL	Greece	gianmosh@statistics.gr, ieromnia@ekt.gr, kleideri@ekt.gr, nmalliou@ekt.gr	National Statistical Service	http://www.statistics.gr/
HU	Hungary	zsuzsanna.szunyogh@ksh.hu	Hungarian Central Statistical Office	https://www.ksh.hu/engstadat
IE	Ireland	Kevin.Phelan@cso.ie, deborah.quinn@djei.ie	Central Statistics Office	http://www.cso.ie/en/index.html
IT	Italy	mastrost@istat.it	National Institute of Statistics	http://en.istat.it/
LV	Latvia	Ilga.Liepina@csb.gov.lv	Central Statistical Bureau	http://www.csb.gov.lv/en
LT	Lithuania	gediminas.samuolis@stat.gov.lt	Official Statistics Portal	http://osp.stat.gov.lt/en/web/guest/home
LU	Luxembourg	Bob.Jung@statec.etat.lu	Statistics Portal	http://www.statistiques.public.lu/en/actors/statec/index.html
MT	Malta	christianne.micallef@gov.mt	National Statistics Office	http://nso.gov.mt/en/Pages/NSO-Home.aspx
NL	Netherlands	h.habets@cbs.nl, k.leufkens@cbs.nl, j.vansteen@rathenau.nl	Statistics Netherlands	http://www.cbs.nl/en-GB/menu/home/default.htm
PL	Poland	D.Rozkrut@stat.gov.pl, M.Mojsiewicz@stat.gov.pl,J.Piotrowska@stat.gov.pl	Central Statistical Office of Poland	http://stat.gov.pl/en/
PT	Portugal	Alexandre.paredes@dgeec.mec.pt	Insituto Nacinal de Estadistica	https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE
RO	Romania	rodica.dumitriu@insse.ro,	Institutul Nacinal de Statistica	http://www.insse.ro/cms/
SK	Slovakia	edita.novotna@statistics.sk	Statistical Office of the SR	www.**statistics**.sk
SI	Slovenia	urska.arsenjuk@gov.si, Darja.Vidmar@gov.si	Statistical Office of the Republic of Slovenia	http://www.stat.si/statweb/en/home
ES	Spain	amaya.saez.alonso@ine.es, miriam.lopez.bahut@ine.es	Instituto Nacional de Estadistica	http://www.ine.es/en/welcome.shtml
SE	Sweden	magnus.ohlson@scb.se, Anna.Sjogren@scb.se	Statistics Sweden	http://www.scb.se/en_/
UK	United Kingdom	cecil.prescott@ons.gsi.gov.uk, jim.nicholls@ons.gsi.gov.uk, hulya.hooker@bis.gsi.gov.uk	Office for National Statistics	http://www.ons.gov.uk/ons/index.html
	EFTA (without Liechtenstein)			
IS	Iceland	Bodvar.Thorisson@hagstofa.is	Statistics Iceland	http://www.statice.is/
NO	Norway	Frank.Foyn@ssb.no, Lars.Wilhelmsen@ssb.no, susanne.sundnes@nifu.no	Statistics Norway	http://www.ssb.no

	Country	R&D	Other statistics	Website of statistical office
CH	Switzerland	Pierre.sollberger@bfs.admin.ch, ELISABETH.PASTOR@bfs.admin.ch	Swiss Federal Statistics Office	http://www.bfs.admin.ch/bfs/portal/en/index.html
	Main Competitors			
US	United States	jjankows@nsf.gov, fmorisor@nsf.gov	United States Census Bureau	http://www.census.gov/
CN	China		National Bureau of Statistics of China	http://www.stats.gov.cn/english/
JP	Japan	tomohiro.ijichi@nistep.go.jp	Statistics Bureau of Japan	http://www.stat.go.jp/english/
KR	South Korea	khdo@kistep.re.kr	Statistics Korea	http://kostat.go.kr/portal/english/index.action
BR	Brazil		Instituto Brasileiro de Geografia e Estadistica	http://www.ibge.gov.br/english/
IN	India		The Statistics Portal	http://www.statista.com/topics/754/india/
IL	Israel	soli@cbs.gov.il, evyatark@cbs.gov.il	The Central Bureau of Statistics	http://www.cbs.gov.il/engindex.htm
CA	Canada	Greg.Peterson@statcan.gc.ca	Canada's National Statistical Agency	http://www.statcan.gc.ca/

Source: Stifterverband

Imputation methods

After data from Eurostat, OECD and national sources have been combined, any remaining gapswill have to be filled by imputation. It has to be decided which imputation method to apply to perform a meaningful analysis. When filling the existing data gaps from 2005 onwards, the following agreements with the EU Commission Services have to be considered:

1. Accuracy of fit: the data gaps may vary in form depending on the country. For instance, whole years or entire industries may be missing over the years. Equally, data gaps may be limited to particular industries or years. The different types of data gaps may require different approaches to fill them. Thus, countries will be grouped according to similarities in the nature of data gaps, with each group being assigned a particular imputation approach.
2. Systematisation: the approach is to be systematised to allow third parties to reproduce it. This includes a prioritisation of approaches in cases where multiple methods are applicable alternatively or complementarily to fill a data gap.
3. Intelligibility: methods that are also communicable to non-statisticians and utilisable for description and induction are to be preferred.

There are two major imputation approaches: single and multiple imputation. The main difference between the two approaches is that the single imputation method fills each missing value with a simulated value, while multiple imputation fills each missing value with a set of values. The statistical complexity and computation load of multiple imputation is significantly higher than single imputation. To decide on the most suitable method, rules of thumb can be used like the specificities of the dataset (continuous vs. ordinal) and the number of missing data compared to the dimension of the dataset. There is, however, no definite answer to the question of the most suitable approach (OECD and JRC, 2008). Imputation applications including the EU KLEMS database and Innovation Union Scoreboard, for example, use single imputation methods. This approach is frequently used because it is considered both simple and efficient. Since our imputation strategy will include single imputation methods, we dedicate the following paragraphs to describing simple single imputation methods we expect to use.

a) Carry forward/backward: backward and forward imputations are among the single imputation methods employed in the aforementioned EU initiatives. These carry values either forwards or backwards along the time series. In the case of consecutive missing values, the carry forward method uses the next preceding non-missing value and the carry backward method the next succeeding non-missing value. Especially at the beginning and end of time series data, this is the most appropriate and transparent approach.

b) Interpolation:

— Replacing missing values with the average of the two values adjacent to the missing value is another very simple imputation method.

— An **auxiliary variable** can be used for the interpolation. For instance, GDP can be used as an auxiliary variable to align national BERD to the economic evolution of the country by implementing a cyclical effect in the imputation. For instance, the Swiss Federal Statistical Office (2013) uses this type of imputation and considers that R&D decisions follow the evolution of the national GDP with a lag of one year rather than simultaneously in order to capture the idea that R&D decisions are made based on economic forecasts. Interpolation is then computed as[25]:

$$BERD_{t_a+i} = BERD_{t_a} \cdot \left(\frac{BERD_{t_b}}{BERD_{t_a}}\right)^{\frac{i}{n}} \cdot \frac{GDP_{t_a+i+1}}{GDP_{t_a+1} \cdot \left(\frac{GDP_{t_b+1}}{GDP_{t_a+1}}\right)^{\frac{i}{n}}}$$

where t_a and t_b are years with non-missing values, with $t_b - t_a = n$, and t_{a+i} is the year for which a value is imputed. This method of using an auxiliary variable can be applied to other variables assuming an appropriate proxy as is the case of GDP for BERD is available and assumptions on the relevance of the time lag are made. The suitability of the auxiliary variable method will be assessed on a case-by-case basis accoounting hence for variable and country specificities.

Example: Switzerland

Available (just total R&D expenditure)

year	2008	2009	2010	2011	2012
R&D expenditure	7,546.9				10,636.3
GDP	376,326.4	388,781.9	439,140.5	501,642.7	518,204.8

Calculation

$$BERD_{2009} = 7{,}546.9 \cdot \left(\frac{10{,}636.3}{7{,}546.9}\right)^{\frac{1}{4}} \cdot \frac{439{,}140.5}{388{,}781.9 \cdot \left(\frac{439{,}140.5}{388{,}781.9}\right)^{\frac{1}{4}}} = 9{,}009.4$$

$$BERD_{2010} = 7{,}546.9 \cdot \left(\frac{10{,}636.3}{7{,}546.9}\right)^{\frac{1}{2}} \cdot \frac{501{,}642.7}{439{,}140.5 \cdot \left(\frac{501{,}642.7}{439{,}140.5}\right)^{\frac{1}{2}}} = 9{,}575.8$$

Additionally, it should be noted that the quality of data ascertained through interpolation depends largely on the period between two reported years. In case of a one-year gap between two reported years, a reliable interpolation can be assumed. The further apart the two observed years are, the less reliable the calculated data becomes**. The limit of tolerance between two observed years is set at four years** (following the approach of the Swiss Federal Statistical Office).

---

25 Another way to look at this imputation is by considering the R&D intensities. The imputed R&D value actually corresponds to an R&D intensity (with the GDP of the following year as the denominator) that is based on the available R&D intensity in t_a, but with an adjustment for both R&D (numerator) and GDP (denominator). $\frac{BERD_{t_a+i}}{GDP_{t_a+i+1}} = \frac{BERD_{t_a}}{GDP_{t_a+1}} \frac{\left(BERD_{t_b}/BERD_{t_a}\right)^{i/n}}{\left(GDP_{t_b+1}/GDP_{t_a+1}\right)^{i/n}}$

c) Ratio imputation: If sectoral data X_t^S for a given year t is not available, while data for both sectoral and a higher aggregated level (NACE 1-digit or country level) for a neighbouring year t_0 are available (respectively $X_{t_0}^S$ and $X_{t_0}^C$), the ratio between the latter figures is multiplied by the current value at the higher aggregated level. This method is appropriate when interpolation cannot be performed. The following equation describes the technique used:

$$X_t^S = \frac{\Sigma X_{t_0}^S}{\Sigma X_{t_0}^C} * \ X_T^C$$

Ratios can be smoothed by computing them over more than one year (see the example below for the United Kingdon).

Example: United Kingdom

For NACE class i: $BERD_i = \frac{\sum_k BERD_{ik}}{\sum_k BERD_k} \cdot BERD_i$ for all $k \neq i$

Available

year	2005	2006	2007	2008	2009
total BERD	13,309.6	14,305.6	15,630.694	15,896.052	15,624
R&D exp. NACE 1.1 class 34			762.118	1,064.208	903.8

Calculation for 2005 and 2006

$$BERD_{34} = \frac{762.118 + 1064.208 + 903.8}{15{,}630.694 + 15{,}896.052 + 15{,}624} \cdot 13{,}309{,}6 = 770.653$$

$$BERD_{34} = \frac{762.118 + 1064.208 + 903.8}{15{,}630.694 + 15{,}896.052 + 15{,}624} \cdot 14{,}305.6 = 828.324$$

d) Differences: in some cases, mainly because of data confidentiality, data are available for a group of sectors while data for one of the sectors is missing. In this case a simple calculation can be used to impute data for the missing sector.

Example: Belgium 2008

Available

NACE 2.0	R&D expenditure
84-85	0.818
84	
85	0.761

Calculation for NACE 84: 0.818 – 0.761 = 0.057

Clearly, country-specific considerations will also be necessary. The extent and type of imputation will be accounted for when interpreting the data, particularly in extreme cases where the great majority of data may have to be imputed.

For variables that are not ratios, we need to ensure that the sum of the sectors is consistent with the total (i.e. country level) figure, if available. Hence, a calibration procedure needs to be implemented in some cases by multiplying the imputed figures by a calibration factor $\theta = X/\hat{X}$, where X is the targeted sum (i.e. the sum we should observe for the imputed sector) and $\hat{X}$ is the sum of the imputed sectors before calibration.

These imputation methods are common approaches among practitioners with the main drawback that they do not reflect the uncertainty of predictions about the unknown missing values. The more advanced single imputation approaches (unconditional mean and regression imputation) do not fully account for imputation uncertainty either.

In particular, when using data involving econometric estimates for research purposes, the following should be noted:

- Configurations: by applying these methods, assumptions or configurations are deliberately integrated into the data (e.g. economic trend-dependent R&D expenditures). The risk when conducting an econometric analysis is that precisely these configurations may emerge as the significant results of the analysis.
- The methods employed may lead to an underestimation of the sampling error, which may in turn have a negative impact on significance tests.

Imputation methods that take this problem into account (although an estimation of the "actual" sampling error should not be assumed) and are recommended by the OECD (e.g. multiple imputation) are typically conceptualised for imputations on a micro-basis and require a suitably large sample of data, unavailable in this case.

Imputation strategy

Although the extent and type of the issues we face depend on the indicators, we follow a harmonised strategy for the imputation of the sectoral data found in Eurostat and OECD. The idea is that identical patterns of missing data will be handled using the same imputation methods. All imputed values are to be documented.

Some missing values can be addressed by simple imputations or by conversion from the available sectoral classifications other than NACE Rev. 2. This strategy takes into account that combining sectors, years and indicators entails vast numbers of individual values which we cannot reasonably expect national offices to provide data for.

The data collected from national sources will be examined in order to ensure comparability. Ordinarily, a complete data set in the NACE 2 classification system will not be achieved by applying only one of the above mentioned methods. Usually, two or more methods will have to be combined, e.g. in the case of incomplete industry data in ISIC. In such a case, completing the data set in ISIC is to take priority over the conversion into NACE.

For the imputation activity, the following priority of methods will be implemented.

1. Converting classifications

   If a complete set of sectoral data is available in another classification (ISIC4, NACE1.1, ISIC3), this set is converted into NACE 2. Due to the revision of the industry classification system NACE 2 in 2008, replacing the previously applicable NACE 1.1, a break may have occurred in time series if countries did not calculate indicators for previous years using the new classification. Provided that the NACE 1.1 data set is complete, the data can be transferred from NACE 1.1 to NACE 2 by means of a concordance table made available by Eurostat[26]. This conversion process will be documented.

   Non-EU member states tend to report on industries not using the European industry classification system NACE, but the international system of industry classification ISIC (with the two-digit level of ISIC 3.1 corresponding to NACE 1.1, and ISIC 4 corresponding to NACE 2 at the 2-digit level). The OECD provides an ISIC/NACE conversion key. ISIC 3.1 data is to be converted first into NACE 1.1, and then into NACE 2 in a second step.

2. Straightforward imputations

   If specific NACE 2 sectors are missing (i.e. one year missing in the time series), missing values are imputed with interpolation (using the variable at country-level as the auxiliary variable for non-ratio indicators) and ratio imputation (using a higher aggregate level).

[26] http://ec.europa.eu/eurostat/web/nace-rev2/correspondence_tables

Interpolations are the first imputation method used if data are missing in-between two years with available data. The maximum number of years to be imputed in this way is 3 years.

If data for the first or last year of a time series are not available, interpolation is not possible. In this case, ratio imputations and carry-forward/backward imputations will be used, depending on the data available. The maximum number of years imputed using this method is 2.

Data provision and updates

Data provision

The construction of a database facilitating the efficient use of all available information is important for current and future usage – and for the production of ID cards. The database will:

1. integrate information from the various data sources,
2. provide input into sector, meta-sector, patent ID cards,
3. be the basis for future data updates.

We decided to use CSV as the file format, because this can be easily transferred to any statistical software or database system (for example MySQL or Oracle SQL as well as Microsoft Access or Microsoft Excel). This database form conforms and complies with the SDMX (Statistical Data and Metadata eXchange) rules and guidelines and their Information Model and with the W3C guidelines for the RDF Data Cube. Essentially, there will be one row in our data sheet for each of the item combinations – in the terminology of SDMX and W3C these are called "dimensions". In our case, these are the year of observation and the country[27] of observation. The data itself is called "measures" in the terminology of the RDF data cube, for which we have one column each. One column for the number of EPO applications in technology field 1, one column for the EPO applications in technology field 2 etc., one column for the number of EPO grants in technology field 1 etc., one column for the number of USPTO applications in technology field 1 etc., and so on. Additional measures arising in the course of the project, for example value added or production, can easily be incorporated by adding columns. Finally, the metadata – called "attributes" in the terminology of the RDF Data Cube – is also stored in individual columns.

An exemplary representation of the database can be found in **Table A13**.

[27] In the case of patents, the regions (NUTS2) are added as "additional countries".

Table A13: Exemplary representation of the database

YEAR	COUNTRY	EPO app Field 1	EPO app Field 2	EPO grant Field 1	USPTO appl	USPTO grant	R&D exp	employ	*Attr EPO	*Attr USPTO	*Attr R&D
2005	AT	200	500	300	150	80	100	200	PATSTAT	PATSTAT	Eurostat
2005	BG	90	30	10	3	2	10	20	PATSTAT	PATSTAT	estimates
2005	CY	25	55	20	5	2	20	80	PATSTAT	PATSTAT	Eurostat
2005	DE	200	2000	800	1000	500	300	400	PATSTAT	PATSTAT	Eurostat
:	:	:	:	:	:	:	:	:	:	:	:
2006	AT	330	550	330	123	21	100	250	PATSTAT	PATSTAT	Eurostat
2006	BG	95	35	20	12	11	10	30	PATSTAT	PATSTAT	estimates
2006	CY	25	65	34	32	30	20	21	PATSTAT	PATSTAT	Eurostat
:	:	:	:	:	:	:	:	:	:	:	:
2007	AT	440	17	9	30	17	15	80	PATSTAT	PATSTAT	Eurostat
2007	BG	100	33	20	10	8	200	30	PATSTAT	PATSTAT	estimates
2007	CY	25	29	11	22	21	300	400	PATSTAT	PATSTAT	Eurostat
2007	DE	250	155	60	75	20	12	20	PATSTAT	PATSTAT	Eurostat
:	:	:	:	:	:	:	:	:	:	:	:

Source: own compilation by Fraunhofer ISI; fictitious data for illustration purposes only.

Specific metadata will be provided that describe the statistical sources (including URLs and hard codes/references) and the specificities of the application including breaks in time series (due to e.g. changing definitions), unavailable data, differences in the definitions between countries, for example, and the algorithms/routines applied to transform/impute the data. The objective is that the metadata contains everything of relevance to understanding the data and needed to replicate the approach.

Updates

Data deliveries are planned in month 12 and month 18. However, updates of the original data sources might not be in accordance with this schedule. For example, the patent database PATSTAT is usually delivered to customers in May/June and October/November. It takes about 3-4 weeks to get the new versions running on our system, so that data production could start in July and December, respectively. Data delivery and updates are planned for December and June. We therefore suggest producing data based on the previous PATSTAT versions so that the respective reports can be produced and then provide data updates (to be entered into the RIO, for example) as soon as these are available. This would be in line with the agreement made during the kick-off meeting that updates will be provided depending on DG Research deliverable timelines and EUROSTAT updates.

In addition, we agreed at the kick-off meeting that updates should include a revision of the data, which includes new data as well as possible revisions in earlier years. So we will always deliver a full data set and not just the most recent years or the latest data in isolation.

Technology definitions

Patents are classified according to their technological content. A very detailed and differentiated classification is used for this purpose: the International Patent Classification (IPC). At the national level, some additional classification systems are in use (e.g. the USPC at the USPTO). In addition, more recently, the Cooperative Patent Classification

(CPC) has been created that will be used in addition to IPC and USPC. However, the database that we use – EPO's PATSTAT – provides information on the IPC classes for all patents, independently of the patent office where it was originally filed. In some cases, automatic conversion/re-classification schemes are used by the EPO to assign IPC classes to each of the patents.

For the purpose of this project, we use three groups of technology definitions, all of which rely on IPC classes. First, we use a definition that covers all patent classes – and therefore all patents – in the database. This is the so called WIPO classification (see **Table A14**), elaborated by Schmoch (2008). The second group deals with Key Enabling Technologies (KETs). Here we also rely on an existing classification provided by the KETs-Observatory (Van de Velde et al. 2013). Finally, we also define the Societal Grand Challenges in terms of technologies. So far – to the best of our knowledge – no technological definition is available. This is why we need to make such a classification ourselves. We describe the classification method used in the following sections.

A general definition of technologies – the WIPO classification

The first classification should cover all patents, be well established and easy to implement. In addition, it should be up to date so that all new IPC classes that are introduced every now and then are covered by this classification scheme.

A number of technology classifications are available, some are more recent, some are updates of earlier classifications, and some are now outdated[28]. Some cover more or less all patent classes that are available, while others only focus on certain technology fields or areas. Several ad-hoc classifications exist that often build on keyword searches. This approach, however, is not suitable for the tasks in this assignment as keyword searches are very time- and resource-consuming and there are a large number of fields to be addressed here. Besides, keyword searches are not stable over time as keywords and fields change in their meaning and especially their relevance for a specific technological field. Finally, and this is the main reason to avoid the approach in this context, it is hard to reproduce keyword searches and use them in large-scale monitoring systems. This is why we suggest focussing on IPC-based classifications and definitions only.

One of the most well-known and widely used classifications is by Schmoch (2008) that was published by WIPO and is used for their annual statistics. It comprises 35 technology classes that are established by the aggregation of several IPC 4-digit classes. It is simple and straightforward to use and is continuously updated – at least as far as the IPC classes are concerned. When new classes occur in the IPC, they are assigned to one of the 35 technology fields. It covers all IPC classes and therefore also all patents. A list of all 35 classes can be found in the following **Table A14**.

[28] One of these outdated lists is the so called OST/INPI/ISI classification that was established in the late 1990s and somehow maintained until the mid 2000s.

Table A14: List of 35 patent classes and their codes according to the WIPO classification

	Area, field	IPC code
I	**Electrical engineering**	
1	Electrical machinery, apparatus, energy	F21H, F21K, F21L, F21S, F21V, F21W, F21Y, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02B, H02G, H02H, H02J, H02K, H02M, H02N, H02P, H05B, H05C, H05F, H99Z
2	Audio-visual technology	G09F, G09G, G11B, H04N 3/%, H04N 5/%, H04N 7/%, H04N 9/%, H04N 11/%, H04N 13/%, H04N 15/%, H04N 17/%, H04N 101/%, H04R, H04S, H05K
3	Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N 1/%, H04Q
4	Digital communication	H04L, H04N 21/%, H04W
5	Basic communication processes	H03B, H03C, H03D, H03F, H03G, H03H, H03J, H03K, H03L, H03M
6	Computer technology	G06C, G06D, G06E, G06F, G06G, G06J, G06K, G06M, G06N, G06T, G10L, G11C
7	IT methods for management	G06Q
8	Semiconductors	H01L
II	**Instruments**	
9	Optics	G02B, G02C, G02F, G03B, G03C, G03D, G03F, G03G, G03H, H01S
10	Measurement	(G01N AND NOT G01N 33/%), G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, G01P, G01Q, G01R, G01S, G01V, G01W, G04B, G04C, G04D, G04F, G04G, G04R, G12B, G99Z
11	Analysis of biological materials	G01N 33/%
12	Control	G05B, G05D, G05F, G07B, G07C, G07D, G07F, G07G, G08B, G08G, G09B, G09C, G09D
13	Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G
III	**Chemistry**	
14	Organic fine chemistry	A61K 8/%, A61Q, C07B, C07C, C07D, C07F, C07H, C07J, C40B
15	Biotechnology	C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S
16	Pharmaceuticals	A61K AND NOT A61K 8/%, A61P
17	Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
18	Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K, C13B 10/%, C13B 20/%, C13B 30/%, C13B 35/%, C13B 40/%, C13B 50/%, C13B 99/%
19	Basic materials chemistry	A01N, A01P, C05B, C05C, C05D, C05F, C05G, C06B, C06C, C06D, C06F, C09B, C09C, C09D, C09F, C09G, C09H, C09J, C09K, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
20	Materials, metallurgy	B22C, B22D, B22F, C01B, C01C, C01D, C01F, C01G, C03C, C04B, C21B, C21C, C21D, C22B, C22C, C22F

	Area, field	IPC code
21	Surface technology, coating	B05C, B05D, B32B, C23C, C23D, C23F, C23G, C25B, C25C, C25D, C25F, C30B
22	Micro-structure and nano-technology	B81B, B81C, B82B, B82Y
23	Chemical engineering	B01B, B01D %, B01D 1%, B01D 2%, B01D 3%, B01D 41/%, B01D 43/%, B01D 57/%, B01D 59/%, B01D 6%, B01D 7%, B01F, B01J, B01L, B02C, B03B, B03C, B03D, B04B, B04C, B05B, B06B, B07B, B07C, B08B, C14C, D06B, D06C, D06L, F25J, F26B, H05H
24	Environmental technology	A62C, B01D 45/%, B01D 46/%, B01D 47/%, B01D 49/%, B01D 50/%, B01D 51/%, B01D 52/%, B01D 53/%, B09B, B09C, B65F, C02F, E01F 8/%, F01N, F23G, F23J, G01T
IV	**Mechanical engineering**	
25	Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66B, B66C, B66D, B66F, B67B, B67C, B67D
26	Machine tools	A62D, B21B, B21C, B21D, B21F, B21G, B21H, B21J, B21K, B21L, B23B, B23C, B23D, B23F, B23G, B23H, B23K, B23P, B23Q, B24B, B24C, B24D, B25B, B25C, B25D, B25F, B25G, B25H, B26B, B26D, B26F, B27B, B27C, B27D, B27F, B27G, B27H, B27J, B27K, B27L, B27M, B27N, B30B
27	Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02B, F02C, F02D, F02F, F02G, F02K, F02M, F02N, F02P, F03B, F03C, F03D, F03G, F03H, F04B, F04C, F04D, F04F, F23R, F99Z, G21B, G21C, G21D, G21F, G21G, G21H, G21J, G21K
28	Textile and paper machines	A41H, A43D, A46D, B31B, B31C, B31D, B31F, B41B, B41C, B41D, B41F, B41G, B41J, B41K, B41L, B41M, B41N, C14B, D01B, D01C, D01D, D01F, D01G, D01H, D02G, D02H, D02J, D03C, D03D, D03J, D04B, D04C, D04G, D04H, D05B, D05C, D06G, D06H, D06J, D06M, D06P, D06Q, D21B, D21C, D21D, D21F, D21G, D21H, D21J, D99Z
29	Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22B, A22C, A23N, A23P, B02B, B28B, B28C, B28D, B29B, B29C, B29D, B29K, B29L, B99Z, C03B, C08J, C12L, C13C, C13G, C13H, F41A, F41B, F41C, F41F, F41G, F41H, F41J, F42B, F42C, F42D, C13B 5/%C13B 15/%C13B 25/%C13B 45/%
30	Thermal processes and apparatus	F22B, F22D, F22G, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24B, F24C, F24D, F24F, F24H, F24J, F25B, F25C, F27B, F27D, F28B, F28C, F28D, F28F, F28G
31	Mechanical elements	F15B, F15C, F15D, F16B, F16C, F16D, F16F, F16G, F16H, F16J, F16K, F16L, F16M, F16N, F16P, F16S, F16T, F17B, F17C, F17D, G05G
32	Transport	B60B, B60C, B60D, B60F, B60G, B60H, B60J, B60K, B60L, B60M, B60N, B60P, B60Q, B60R, B60S, B60T, B60V, B60W, B61B, B61C, B61D, B61F, B61G, B61H, B61J, B61K, B61L, B62B, B62C, B62D, B62H, B62J, B62K, B62L, B62M, B63B, B63C, B63G, B63H, B63J, B64B, B64C, B64D, B64F, B64G
V	**Other fields**	
33	Furniture, games	A47B, A47C, A47D, A47F, A47G, A47H, A47J, A47K, A47L, A63B, A63C, A63D, A63F, A63G, A63H, A63J, A63K
34	Other consumer goods	A24B, A24C, A24D, A24F, A41B, A41C, A41D, A41F, A41G, A42B, A42C, A43B, A43C, A44B, A44C, A45B, A45C, A45D, A45F, A46B, A62B, A99Z, B42B, B42C, B42D, B42F, B43K, B43L, B43M, B44B, B44C, B44D, B44F, B68B, B68C, B68F, B68G, D04D, D06F, D06N,

	Area, field	IPC code
		D07B, F25D, G10B, G10C, G10D, G10F, G10G, G10H, G10K
35	Civil engineering	E01B, E01C, E01D, E01F 1/%, E01F 3/%, E01F 5/%, E01F 7/%, E01F 9/%, E01F 11/%, E01F 13/%, E01F 15/%, E01H, E02B, E02C, E02D, E02F, E03B, E03C, E03D, E03F, E04B, E04C, E04D, E04F, E04G, E04H, E05B, E05C, E05D, E05F, E05G, E06B, E06C, E21B, E21C, E21D, E21F, E99Z

Note: This table is available in Excel format at: http://www.wipo.int/ipstats/en/statistics/technology_concordance.html

Source: WIPO IPC-Technology Concordance Table, Schmoch 2008

KETs

Key Enabling Technologies are defined as industrial biotechnology, nano-technology, micro- and nano-electronics, photonics, advanced materials, and advanced manufacturing technologies (AMT). These technologies are assumed to have a cross-cutting and enabling character, and to be relevant in a number of other fields and areas.

The KETs Observatory (Aschhoff et al. 2010; Van de Velde et al. 2013) has suggested definitions and classifications of these key enabling technologies, in terms of both sectors (NACE) and technologies (IPC)[29]. A list of IPC classes for each of the key enabling technologies can be found in **Table A15**. A matrix to calculate economic indicators on the level of KETs was also provided, so that all relevant indicators can be directly calculated. However, when it comes to R&D data, the heterogeneity of companies and sectors needs to be taken into account by introducing technology weights which are capable of capturing the different R&D intensities. We therefore suggest recalculating the R&D data (available on the level of sectors) using our method of estimation (see section 5 of this report), instead of using the direct sectors provided by the KETs Observatory.

Table A15: IPC classes of Key Enabling Technologies

KETs	IPC classes
Nano-technology	B82Y (previously Y01N), B81C, B82B
Photonics	F21K, F21V, F21Y, G01D 5/26, G01D 5/58, G01D 15/14, G01G 23/32, G01J, G01L 1/24, G01L 3/08, G01L 11/02, G01L 23/06, G01M 11, G01P 3/36, G01P 3/38, G01P 3/68, G01P 5/26, G01Q 20/02, G01Q 30/02, G01Q 60/06, G01Q 60/18, G01R 15/22, G01R 15/24, G01R 23/17, G01R 31/308, G01R 33/032, G01R 33/26, G01S 7/481, G01V 8, G02B 5, G02B 6 (excl. subclasses 1, 3, 6/36, 6/38, 6/40, 6/44, 6/46), G02B 13/14, G03B 42, G03G 21/08, G06E, G06F 3/042, G06K 9/58, G06K 9/74, G06N 3/067, G08B 13/186, G08C 19/36, G08C 23/04, G08C 23/06, G08G 1/04, G11B 7/12, G11B 7/125, , G11B 7/13, , G11B 7/135, G11B 11/03, G11B 11/12, G11B 11/18, G11C 11/42, G11C 13/04, G11C 19/30, H01J 3, H01J 5/16, H01J 29/46, H01J 29/82, H01J 29/89, H01J 31/50, H01J 37/04, H01J 37/05, H01J 49/04, H01J 49/06, H01L 31/052, H01L 31/055, H01L 31/10, H01L 33/06, H01L 33/08, H01L 33/10, H01L 33/18, H01L 51/50, H01L 51/52, H01S 3, H01S 5, H02N 6, H05B 33
Industrial bio-technology	C02F 3/34, C07C 29, C07D 475, C07K 2, C08B 3, C08B 7, C08H 1, C08L 89, C09D 11, C09D 189, C09J 189, C12M, C12P, C12Q, C12S, G01N 27/327 except for co-occurrence with A01, A61, C07K 14/435, C07K 14/47, C07K 14/705, C07K 16/18, C07K 16/28, C12N 15/09, C12N 15/11, C12N 15/12, C12N 5/10, C12P 21/08, C12Q 1/68, G01N 33/15, G01N 33/50, G01N 33/53, G01N 33/68, G01N 33/566, C12N 1/19, C12N 1/21, C12N 1/15, C12N 15/00, C12N 15/10, C12P 21/02.
Advanced materials	B32B 9, B32B 15, B32B 17, B32B 18, B32B 19, B32B 25, B32B 27, B82Y 30, C01B 31, C01D 15, C01D 17, C01F 13, C01F 15, C01F 17, C03C, C04B 35, C08F,

[29] https://webgate.ec.europa.eu/ketsobservatory/library

KETs	IPC classes
	C08J 5, C08L, C22C, C23C, D21H 17, G02B 1, H01B 3, H01F 1/0, H01F 1/12, H01F 1/34, H01F 1/42, H01F 1/44, H01L 51/30, H01L 51/46, H01L 51/54.
Micro- and nanoelectronics	G01R 31/26, G01R 31/27 , G01R 31/28 , G01R 31/303 , G01R 31/304, G01R 31/317, G01R 31/327, G09G 3/14, G09G 3/32, H01F 1/40, H01F 10/193, H01G 9/028, H01G 9/032, H01H 47/32, H01H 57, H01S 5, H01L, H03B 5/32, H03C 3/22, H03F 3/04, H03F 3/06, H03F 3/08, H03F 3/10, H03F 3/12, H03F 3/14, H03F 3/16, H03F 3/183, H03F 3/21, H03F 3/343, H03F 3/387, H03F 3/55, H03K 17/72, H05K 1, B82Y 25 (certain overlap to nanotechnology).
Advanced Manufacturing Technologies	B01D 15, B01D 67, B01J 10, B01J 12, B01J 13, B01J 14, B01J 15, B01J 16, B01J 19/02, B01J 19/08, B01J 19/18, B01J 19/20, B01J 19/22, B01J 19/24, B01J 19/26, B01J 19/28, B01J 20/30, B01J 21/20, B01J 23/90, B01J 23/92, B01J 23/94, B01J 23/96, B01J 25/04, B01J 27/28, B01J 27/30, B01J 27/32, B01J 29/90, B01J 31/40, B01J 38, B01J 39/26, B01J 41/20, B01J 47, B01J 49, B01J 8/06, B01J 8/14, B01J 8/24, B01J 10, B01L , B04B , B04C , B32B 37, B32B 38, B32B 39, B32B 41, B81C 3, B82B 3, B82Y 35, B82Y 40, C01B 17/20, C01B 17/62, C01B 17/80, C01B 17/96, C01B 21/28, C01B 21/32, C01B 21/48, C01B 25/232, C01B 31/24, C01B 9, C01C 1/28, C01D 1/28, C01D 3/14, C01D 5/16, C01D 7/22, C01D 9/16, C01F 1, C01G 1, C02F 11/02, C02F 11/04, C02F 3, C03B 20, C03B 5/24, C03B 5/173, C03B 5/237, C03B 5/02, C03C 21, , C03C 29, C04B 11/028, C04B 35/622, C04B 35/624, C04B 35/626, C04B 35/653, C04B 35/657, C04B 37, C04B 38/02, C04B 38/10, C04B 40, C04B 7/60, C04B 9/20, C07C 17/38, C07C 2/08, C07C 2/46, C07C 2/52, C07C 2/58, C07C 2/80, C07C 201/16, C07C 209/82, C07C 213/10, C07C 227/38, C07C 231/22, C07C 249/14, C07C 253/32, C07C 263/18, C07C 269/08, C07C 273/14, C07C 277/06, C07C 29/74, C07C 303/42, C07C 315/06, C07C 319/26, C07C 37/68, C07C 4/04, C07C 4/06, C07C 4/16, C07C 4/18, C07C 41/34, C07C 41/58, C07C 45/78, C07C 45/90, C07C 46/10, C07C 47/058, C07C 47/09, C07C 5/333, C07C 5/41, C07C 51/42, C07C 51/573, C07C 51/64, C07C 57/07, C07C 67/48, C07C 68/08, C07C 7, C07D 201/16, C07D 209/84, C07D 213/803, C07D 251/62, C07D 301/32, C07D 311/40, C07D 499/18, C07D 501/12, C07F 7/20, C07H 1/06, C07K 1, C08B 1/10, C08B 17, C08B 30/16, C08C , C08F 2/01, , C09B 41, C09B 67/54, C09D 7/14, C09J 5, C12M, C12S , C21C 5/52, C21C 5/54, C21C 5/56, C21C 7, C21D , C22B 11, C22B 21, C22B 26, C22B 4, C22B 59, C22B 9, C22C 1, C22C 3, C22C 33, C22C 35, C22C 47, C22F , C23C 14/56, C23C 16/54, C25B 9, C25B 15/02, C25C , C25D 1, C30B 15/20, C30B 35, C40B 60, D01D 10, D01D 11, D01D 13, D01F 9/133, D01F 9/32, D06B 23/20, D21H 23/20, D21H 23/70, D21H 23/74, D21H 23/78, D21H 27/22, F24J 1, F25J 3, F25J 5, F27B 17, F27B 19, F27D 19, F27D 7/06, G01C 19/5628, G01C 19/5663, G01C 19/5769, G01C 25, G01R 3, G11B 7/22, H01L 21, H01L 31/18, H01L 35/34, H01L 39/24, H01L 41/22, H01L 43/12, H01L 51/40, H01L 51/48, H01L 51/56, H01S 3/08, H01S 3/09, H01S 5/04, H01S 5/06, H01S 5/10, H05B 33/10, H05K 13, H05K 3

Source: KETs Observatory (van de Velde et al. 2013).

Societal Grand Challenges

A shift in terms of application orientation is envisaged with Horizon 2020 (H2020) as part of the Innovation Strategy in European policy making. This implies a focus on pathways to technological solutions. Previous thematic priorities and research lines under FP6 and FP7 have been retained – the main change within H2020 is not the thematic priorities, but the type of projects funded. In an attempt to be less prescriptive, the Work Programmes (WP) under the Grand Challenges – which encompass previous thematic priorities to a large extent – try to indicate the areas expected to be targeted. As mentioned above, the type of projects funded under H2020 differ; in a number of specific areas, we can thus find up to four dedicated call texts addressing more basic research questions, applied ones (up to demonstration level), a policy level call and a diffusion call. In other areas, there are two stage calls: the first aims at development and the second at diffusion/commercialisation. All this offers more flexibility and freedom than under FP7, but in terms of definitions and analyses, this openness is too broad or too unspecific to be operationalized and measured. There is no clear definition of what should be covered by the Societal Grand Challenges (SGCs) in terms of included or excluded scientific or technological fields. Furthermore, there are no definitions or demarcations for SGCs in terms of patent classes.

The following seven Grand Challenges are to be addressed in the course of this assignment:

— Health, demographic change and well-being,
— Food security, sustainable agriculture, marine and maritime research, and the bioeconomy,
— Secure, clean and efficient energy,
— Smart, green and integrated transport,
— Climate action, resource efficiency and raw materials,
— Secure societies – protecting the freedom and security of Europe and its citizens,
— Europe in a changing world – inclusive, innovative and reflective societies.

While we are able to rely on existing technological definitions and demarcations in the case of overall technologies and KETs, this is not the case for SGCs. As a consequence, we need to define SGCs in terms of IPC classes. During the kick-off meeting and the discussions during the project so far, we agreed on the following procedure to achieve a technological definition of SGCs. Given that the definition may become publicly available and might spill over into other policy analyses, statistical examinations, and scientific exercises, there is extra pressure on defining the SGCs properly.

As a starting point, we held a workshop in Brussels at the end of March with a number of experts on technology classifications, data treatment, and statistical analyses from a broad array of institutions like academia, the EPO, EUROSTAT, IPTS as well as other policy makers from the Commission Services. Two main results emerged from the workshop. First, there is continuity in policy making that makes it possible to draw on previous analyses and findings, for example, in the context of FP7. In addition, a modular definition of technologies within SGCs allows a sufficient level of flexibility in adapting the definitions. Second, the following procedure was suggested to elaborate a technology definition:

— Check the (current and maybe upcoming) work programme and legal basis of H2020 for funded technologies;
— Set up interviews and discussions with representatives for each SGC from DG RTD, DG CONNECT and DG HOME to grasp their understanding of the SGCs;
— Use the information gathered to suggest technological definitions that are then re-discussed with the representatives; rely on existing classifications as far as possible.

In the course of these discussions with representatives, it became clear that it was also worthwhile to examine the newer WP to obtain a more general and longer-term perspective of each SGC rather than only the current WP. This was taken into account when fine-tuning and adapting the technology definitions. In addition, during the structured interviews and discussions, during which first general technology lists were provided by the contractors (based on the available WPs of the SGC), a common understanding of the SGCs was achieved, and all the representatives agreed to check the precise technology definitions suggested by the contractors after the interviews. All the representatives agreed that the view of SGCs should be broader and more general than in the current WP under H2020. In addition, several SGCs address general questions some of which might not be directly funded under H2020, but through a Joint Undertaking (JU) like the Innovative Medicines Initiative (IMI) as well as private funding and research by companies. The latter is particularly important here as the core of this project deals with business R&D (BERD) and patent applications, about 90% of which originate in the private sector. So our definition of the SGCs is in the broad sense of the challenges and not in the narrower sense of the funding programme.

As far as possible, we rely on existing patent classifications and assign them (totally or in parts) to the respective SGC rather than creating completely new ones.

For example, the OECD[30] provides definitions of selected fields, namely biotechnology, ICT, nano-technology and environmental technologies. Furthermore, several patent offices publish search strategies and technology definitions for particular fields in their annual reports, for example, electric and hybrid vehicle technologies, renewable energy technologies, or biotechnology.[31]

The EPO used to classify nano-technologies (Y01), climate mitigation technologies (Y02) and more recently also energy-saving technologies in smart grids (Y04) using their internal classification scheme, the ECLA (European Classification System) that supplemented the IPC. The ECLA was then transferred to the CPC (Cooperative Patent Classification), which is jointly managed by the EPO and the USPTO. Y02 and Y04 technologies are still classified there[32], while Y01 was integrated into the IPC in 2010 and is now classified as B82. These technology definitions are also available in PATSTAT. The SETIS project as well as an EPO/UN report on Climate Chance and Mitigation Technologies (CCMT) uses this classification so that it is well established in the field and accepted by stakeholders in the Commission Services as well as outside the Commission.

In a project for DG-RTD called "Measurement and analysis of knowledge and R&D exploitation flows, assessed by patent and licensing data", INCENTIM at K.U. Leuven together with KITES at the University Bocconi developed a concordance between FP7 thematic priorities and IPC classes (Patent Indicators by Thematic Priority)[33]. They based their technological definitions on Schmoch (2008) and assigned them to 15 thematic priorities. As they also kept a differentiation below the level of these 15 priorities, a modular definition of the thematic priorities was achieved, which can also be used in this project to assign technologies to one of the SGCs.

These available classification schemes form the basis for our definitions of the SGCs.

Conversion of data

The main aims of the study are to analyse the performance of the EU and individual EU member countries over time and in comparison to associated and competing countries on the sectoral level, at the level of Key Enabling Technologies (KETs) and with respect to the Societal Grand Challenges (SGCs). To achieve these aims, the most crucial task within this study is to process data from different sources and especially with different demarcations/classifications in such a way that they can be jointly analysed. To be more precise, the biggest challenge is to shift patents to the sectoral level and, vice versa, to shift the data available at sector level, e.g. BERD, to the level of technologies.

Patents by sectors – from technologies to NACE 2

The first challenge is to provide patent data by sector. According to Schmoch (2008), 'sectors' and 'technologies' describe different aspects of products and need to be treated differently.

---

30 Please refer to http://www.oecd.org/innovation/inno/oecdworkonpatentstatistics.htm.

31 See for example the annual reports by the German Patent and Trademark Office at http://www.dpma.de/english/service/publications/annualreports/index.html.

32 In the call for this tender, this was mentioned as reference to the "IPC Green Inventory".

33 Annelies Geerts, Gianluca Tarasconi, Francesca Innocenti, Xiaoyan Song, Julie Callaert, Maikel Pellens, Caro Vereyen, Cathy Lecocq, Stefano Breschi, Bart Van Looy (2011): Measurement and analysis of knowledge and R&D exploitation flows, assessed by patent and licensing data. Deliverable 1.6: Patent Indicators by Thematic Priority; Leuven, Milano.

Existing concordance schemes

There have been a number of suggestions in the past on how to merge technological and sectoral classifications. One of the most well known and widely used approaches was suggested by Schmoch et al (2003). They used a microdata approach to match 44 technological fields to 43 industrial sectors, providing a concordance matrix to assign patent counts to sectors. The basic idea was to collect patent data according to the 4-digit IPC classes for the 44 technological fields, then apply the matrix, which contains the shares of sectors per technological field. Unfortunately, the approach has since been misunderstood or misused by several users, most recently by Lybbert and Zolas, for example (Lybbert, Zolas 2012). Instead of applying the concordance matrix, they used the 44 technology classes directly and assigned them to sectors. The original procedure suggested by Schmoch et al. (2003) is rather complex, but it was empirically based and had proved its reliability in empirical tests.

This concordance matrix, however, has certain drawbacks due to the fact that it was established about 12 years ago. First of all, it uses IPC7 as the basis for the technological classification – a classification scheme that is no longer being used. The new IPC scheme differs considerably in certain relevant aspects, for example, in the use of main and secondary classes and in the coverage of certain new and emerging technologies. This latter aspect is of particular importance in our context as the Key Enabling Technologies and some parts of the Societal Grand Challenges fall into this group. Furthermore, a number of new countries have since become actors on the R&D stage and the R&D expenditure in most countries has not only increased in absolute (nominal), but also in relative terms. This means that the relation between sectors and technologies as empirically defined by Schmoch et al. (2003) is most probably no longer accurate.

As many users were not able or willing to apply the concordance matrix, but tended to use the shortcut of directly assigning technologies to sectors, Schmoch and Gauch (2004) then suggested a simplified version of the 44 technological fields used in the 2003 study and assigned them directly to sectors. They provided a list of 19 sectors and the corresponding IPC 4-digit classes. This was a purely intellectual not an empirical concordance and it made a 1:1 assignment of individual IPC classes to individual sectors, rather than assigning a probability (or fractional) as was the case in Schmoch et al. (2003).

In 2002, Johnson (Johnson 2002) suggested a concordance based on the intellectual assignments of SIC codes by patent examiners at the Canadian patent office. He also applied a probabilistic approach for his concordance. The interesting feature was that the patent examiners assigned a sector of invention as well as a sector of use to each patent they examined. However, the Canadian patent office stopped this extra work in the 1990s and therefore no new data was available. Johnson (2002), however, used the data up to the 1990s to develop a concordance between IPC and SIC/ISIC sectors, also differentiating by sector of invention and sector of use. The OECD, who funded the study, used this approach for their statistics and patent analyses in the first half of the 2000s. Subsequently, they also switched to the concordance suggested by Schmoch et al. (2003).

There were a number of other approaches (see, for example, Evenson, Putnam 1988; Verspagen et al. 1994) in the 1990s and even 1980s using different methods, but most of them employed intellectual assignments of IPC classes to sectors – either made by the researchers themselves or also those made by the Canadian patent examiners.

More recently, another approach was suggested by Lybbert and Zolas (Lybbert, Zolas 2012), who argued that direct (100%) assignments of IPC classes to sectors are not adequate as neither companies nor sectors are technologically homogenous. This is an argument already made by Schmoch et al. (2003). As already pointed out, their concordance was not a deterministic but a probabilistic assignment, as long as the procedure was followed as intended by the authors. Lybbert and Zolas, however, used a keyword-based algorithm to make a probabilistic matching of patents to sectors. Their procedure is rather complex and time consuming. Moreoever, the stability of the results/assignments and the procedure's applicability to new and emerging fields, where the wording and use of terms are in flux, still need to be demonstrated.

Finally, most recently, van Looy, Vereyen and Schmoch (van Looy et al. 2014) provided an updated version of the concordance established in 2003 by Schmoch et al. in a publication edited by Eurostat, which intends to apply it to its patent statistics. Eurostat was one of the most intensive and long-term users of the Schmoch et al. (2003) concordance, but the above mentioned shortcomings made an update (or alternative) necessary. Van Looy et al. (2014) updated the 44 technology definitions and checked the assignments and groupings of each of the IPC 4-digit codes. In particular, they were able to take the new NACE 2 classification into account, while previous concordances had to rely on NACE 1.1. However, their work also resulted in a direct 1:1 assignment of individual IPC classes to individual sectors and is not – like the approaches in Schmoch et al. (2003) or Lybbert and Zolas (2012)– a probabilistic concordance. It is therefore easy to implement, for example in PATSTAT, but faces the same shortcomings as any other direct assignment, namely that it is not able to take the heterogeneity of sectors in terms of technologies into account. Furthermore, it is not a direct, empirically-derived concordance, but mainly an intellectual concordance that uses empirical data for plausibility and distribution checks.

Setting up a new probabilistic concordance scheme

As there is no probabilistic concordance scheme that takes into account the current industrial structures – except the one suggested by Lybbert and Zolas (2012), which has the shortcomings mentioned above – we will use the Schmoch et al (2003) approach to set up a new probabilistic concordance scheme. This has become feasible as data availability has improved in the past decade and the formerly difficult task of providing microdata on the sector for each of the patent applicants is now easier to achieve.

We therefore suggest establishing a more topical concordance. To achieve this goal, we will use a matched data set of PATSTAT with BvD's Orbis. The microdata set will contain information on the sector of each applicant as well as the numbers and shares of patents by technological fields of each company (technological profile). Aggregated across all companies to the level of sectors, the shares of each particular technology field can be assigned to each of the sectors. In more detail, the working steps involved in setting up the new probabilistic concordance scheme are as follows. First of all, we use the technology classification of WIPO (Schmoch 2008) to define the technologies. Secondly, we collect the most recent patent data for each of the companies. Thirdly, we aggregate the patent data on the level of individual sectors. Finally, the shares of each of the sectors per technology will be calculated, which then represents the probabilistic distribution of the new concordance matrix. This matrix can then be applied to any patent data collected using the WIPO classification. If, for example, sector A is responsible for 50% of the patents in technology field x and sector B is responsible for 30%, and sector C for 20%, the number of patents is split accordingly into these three sectors. Concerning technology y, the shares of the sectors are 20%, 20%, and 60%, respectively, and the patents in this technology field are assigned according to this distribution. The number of patents in sector A (or B or C) is then simply the sum of patents across all technologies. This is exactly the same procedure that was used by Schmoch et al (2003), but it reflects the changes in company profiles as well as in technological classification over the last 11 years.

However, unlike Schmoch et al. (2003), we will also check whether it is feasible and reasonable to employ different concordances/matrixes for different countries or groups of countries – for example one for western industrialised countries, one for Southern European countries, one for Asian countries, and one for the rest of the world. Finally, we will compare our concordance scheme with the one by van Looy et al. (2014) and discuss the advantages and disadvantages of each with Eurostat representatives and other researchers. Finally, there will be a joint decision with IPTS and Eurostat representatives on which of the two concordance schemes we will apply for the data collection in this project.

R&D by technologies – from sectors to technology groups

For the sector analysis (see previous section), it is necessary to convert the patent data (technologies) into sectoral data. To be able to provide R&D data (and other economic data) by technology, a reverse conversion is necessary. The project plans to provide this data for the Key Enabling Technologies and the Societal Grand Challenges.

What, at first sight, may seem to be a simple problem, similar to the conversion of patents into sectors, becomes a trickier endeavour when looked at more closely. While a direct link of sectors and technologies exists via the applicant, who files patents for certain technologies and is assigned a certain sector (based on BvD's Orbis data), this link is only indirectly available when converting from sectors to R&D as the distribution of R&D by technologies is not directly available. In other words, we know which companies from which sector are filing patents for which kind of technologies, but we do not know how much R&D each company in a particular sector spends on R&D for which kind of technologies.

The method to convert R&D by sector into R&D by technology is described in this section. Again, we use a microdata approach, but in this case we match the EU R&D Scoreboard with PATSTAT as we need a data set that not only comprises sector and technology information, but also data on R&D expenditure. Starting from the definitions of KETs and SGCs (see above), the number of patents in each field will be calculated for each of the companies in the matched dataset. Next, we calculate the share of these patents in total patents by sector. This gives us the share of KETs and SGCs for each of the sectors. In principle this would allow us to use these shares – under the assumption that any economic indicator follows the same distribution as the patents – to calculate the economic indicators. For example, if 5% of all patents in the sector are KETs, we simply calculate that 5% of the employment, the R&D expenditure or the turnover can also be assigned to the technological fields of KETs. This is exactly what was suggested in the feasibility study of the KETs Observatory (Van de Velde et al. 2013). While this might be a reasonable assumption for employment, maybe less so for turnover, this is absolutely not the case concerning R&D expenditures or R&D personnel. The reason is that patents require different amounts of R&D investment. In other words, different technologies have different R&D intensities (as well as different patent intensities). As neither companies nor sectors are homogeneous in their technological activity – and where emerging technologies are concerned, it is justified to assume an even more scattered distribution across sectors and companies – they also have different R&D intensities.

A recently completed project of Fraunhofer ISI and Stifterverband on behalf of the German Federal Ministry of Education and Research re-calculated the R&D expenditure by technologies (Frietsch et al. 2014). This study matched the German R&D survey with the PATSTAT database. In principle, this result/matrix could also be employed in this project. However, as the concordance was constructed based only on German data, it might not be directly applicable to other countries. This is why we suggest using the same approach, but applying it to the R&D Scoreboard data matched with PATSTAT as it was performed in a recent project on behalf of IPTS (Gkotsis 2015; Neuhäusler et al. 2015).

In more detail, the following steps will be implemented. First, we calculate the share of each individual KET/SGC in total patents per company, assuming that the patents are an indication of the whole technological profile of that particular company. It is not relevant for this approach that patents represent 100% of the technological activities of each individual company, but it is important that the patent profiles reflect the technological profiles of the respective company. In other words, it is not important that a company patents all its inventions, but it is important that the relation between patented and total inventions is similar across all the technological fields in which this particular company is active. However, we cannot control or check this, so it has to remain an assumption.

In a second step, we aggregate the number of patents in the particular KET/SGC per sector and calculate the share of that particular technology field (KET/SGC) for each of the sectors. We could apply this new matrix directly to the economic data for the re-calculation. For example, if sector A has 1000 employees and the share of KET/SGC x in sector A is 10%, then we calculate that 100 employees in sector A are working in KET/SGC x. We do this for all the sectors in our sector classification and then sum up the number of employees across all sectors to find out how many people are employed in KET/SGC x.

However, there is one issue that needs to be taken into account, which holds for any economic data, but which is most obvious in terms of R&D expenditure and R&D personnel. Neither companies nor sectors are homogeneous in their technological profiles. As technologies differ in their R&D requirements (R&D intensity), we cannot assume that the same amount of R&D expenditure will produce the same number of patents in technology x as in technology y. For example, a pharmaceutical patent may cost (in terms of R&D) 1 million euros and a household appliance patent €100,000. If a company active in these two technological fields spends 5.5 million euros and holds five patents in pharmaceuticals and five patents in household appliances, we cannot assume that 50% of its R&D is spent on pharmaceuticals and the other 50% on household appliances. It is more likely to be the case that the company spends 5 million euros on pharmaceuticals and €500,000 on household appliances.

So what we need to take into account is the R&D intensity per patent in each field. In other words, we need to develop a weighted matrix rather than an unweighted one. The weights will be derived from external sources, as otherwise there would be the danger of creating a tautology. On the macro level, we calculate the R&D intensity per patent based on the intellectual sector-technology concordance suggested by Schmoch and Gauch (2004). The sector-industry link will be used to calculate the R&D intensity per technology using BERD data by sector for the average of all OECD countries. This R&D intensity will be applied to the matched microdata of the EU R&D Scoreboard and PATSTAT, thereby assigning a higher weighting to patents in more expensive (in terms of R&D) technology fields and a lower one to patents in less expensive technology fields. Again, we aggregate them for each of the sectors and calculate the share of each KET/SGC within each sector. What we then obtain is a weighted concordance matrix that we can apply directly to the R&D data by sector.

To sum up, we will collect the data for KETs and SGCs for each of the companies in the dataset. Then, and this is the crucial additional step, we will use the R&D intensity (R&D expenditure per patent) by technology group to weight the patents in each company. More expensive patents will have higher weights, less expensive patents lower weights. The R&D expenditure will be derived based on the simple sector-technology concordance provided by Schmoch and Gauch (2004). Next, we calculate the share of each technology in total patents for each of the sectors. The result is a single vector with shares of individual technologies in sectors, which can then be used to calculate the amount of R&D expenditure in each sector on each particular technology by applying it to BERD data by sector. So the matrix is established/created using micro-level data, but is then applicable to the macro/meso-level data of R&D expenditure by sector.

We will have to see whether we again create particular matrices for certain country groups – for example one for western industrialised countries, one for Southern European countries, one for Asian countries, and one for the rest of the world. In addition, we will check whether one matrix can be applied for the conversion of all economic data – R&D expenditure, turnover, value added – or if we need to establish specific matrices for each economic indicator, which could be done by using weighting schemes instead of the R&D intensity derived from applying the concordance suggested by Schmoch and Gauch (2004).

Patent data collection

With respect to patent data, the task is to collect patents using EPO's PATSTAT database for each of the sectors, the individual KETs and individual SGCs as well as for the sectors, KETs and SGCs in total for the period starting in 2000 to the most recently available year. We will assign patents to years employing the priority year information instead of the application date (or even the publication date) as this is the first date on a patent and is independent of the processing at the particular office and the filing strategy of the particular applicant. In addition, and much more important here, it is the closest date available in patent documents to the date of invention or at least to the finalisation of the invention process. If patents are used as an output indicator of R&D processes, then this is the best choice, as one can expect that it is closest to the "event" of R&D expenditure. It needs to be stressed, however, that due to the legal obligations of the patent system, patents are only published 18 months after the priority date. For our statistics and analyses, this has the consequence that we have a two year delay in the availability of the data. For example, the latest comprehensive available year for the PATSTAT release of September 2015 will be 2013. Of course, we could also use the publication date and then provide data up to the year 2014, but this is only window-dressing as it is essentially the same underlying data that simply appears more topical, but at the cost of less "compatibility" with the R&D data due to the reasons mentioned above.

First of all, this holds for data on patent applications. When it comes to granted patents, the processing time at the patent office has a strong impact. We suggest that granted patents should also be collected according to the priority date, again, as this is assumed to be the closest date to the R&D event. The average processing time of a patent is about 3 to 4 years – as claimed by the EPO[34] or the USPTO[35], for example. However, it takes about 7 to 8 years after priority filing until about 90% of the patents of each cohort are processed. This means that with the September 2015 release of PATSTAT, one can expect to have fairly complete data for the priority years 2008 or 2009.

We will provide the data for a large set of countries, namely: Austria (AT), Belgium (BE), Brazil (BR), Bulgaria (BG), Canada (CA), Switzerland (CH), China (People's Republic) (CN), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Spain (ES), Finland (FI), France (FR), Croatia (HR), Hungary (HU), Ireland (IE), Israel (IL), India (IN), Iceland (IS), Italy (IT), Japan (JP), South Korea (KR), Lithuania (LT), Luxembourg (LU), Latvia (LV), Malta (MT), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Sweden (SE), Slovenia (SI), Slovakia (SK), United Kingdom (UK), and United States of America (US). In addition, country aggregates like EU-28, EFTA, FERA, BRIC and OECD will be provided separately as the totals of these country groups might deviate from the sum of the countries, because we apply whole counting instead of fractional counting. The patents will be assigned to countries based on both the applicant's address listed on the patent and the inventor's address. We use whole counts, which means if there is more than one applicant from different countries, the patent will be fully assigned to each of the countries. This means that such patents are double counted. In effect, the sum of all patents across all countries will be higher than the total number of patents filed at the respective office in the particular year.

In addition, we also provide data on fractional counting by countries, which means that each patent is assigned only partially to the country of inventor. The fraction is defined by the number of inventors from each particular country. For example, two French, one Italian and one Belgian inventor filing one patent would result in ½ of the patent counted for France, and ¼ each for Italy and Belgium.

[34] See: http://www.epo.org/about-us/annual-reports-statistics/annual-report/2014/statistics/ quality-indicators.html; the time to grant after the request for examination is reported as being 26 months on average. However, the request for examination is possible up to about 2 years after priority filing – even later in the case of PCT patents.

[35] See http://www.uspto.gov/about/stratplan/ar/USPTOFY2014PAR.pdf, p. 47/48; the USPTO provides data on the duration from filing to the decision or withdrawal. The current time is about 27 months, which is considerably faster than the EPO – at least on average.

We will provide the following patent data for each country and country aggregate (inventor and applicant)[36], each year and each sector/field as well as the totals:

1. Number of EPO, USPTO and PCT patent applications and grants,
2. Number of EPO, USPTO and PCT patent applications and grants by type of applicant (public, private, individual),
3. Number of international co-patents at EPO (applications and grants) and via the PCT procedure (whole country only!),
4. Number of co-patents between different types of research performers[37] at EPO (applications and grants) and via the PCT procedure. This is essentially the number of patents with at least two different applicants of two different types, independent of their country (whole counting only!),
5. Number of triadic[38], Transnational[39], or IP5[40] patent applications,
6. Number of highly cited[41] patents (at the EPO only!),
7. Number of patents citing non-patent literature (at the EPO only!),
8. A consolidated list of applicants active in KETs and SGCs by country in a particular period.

Sector/technology ID cards

For each country and each technology field, so called ID cards are required, which not only provide a selection of tables and graphs, but also brief descriptions of the trends and results. Such an ID card needs to be produced in the first year of the contract and updated in the second year for 39 countries (and 5 additional country groups), 6 KETs and 7 SGCs as well as 35 technological fields (WIPO classes). Given the large number of cards involved, we decided to automatically produce them. They can easily be updated whenever new data is available. The ID cards are a stand-alone output of this study, but also form the annex for the three cover reports on sectors, KETs and SGCs. Each ID card will comprise tables and graphs of the relevant data as well as the interpretation of trends and explanatory elements which will summarise the most relevant aspects for each sector/field. For this purpose, we will develop a script (in SPSS) that produces sectoral ID cards and a Java script that automatically produces text, interpreting the results and trends in the sectoral ID cards. Trends, growth rates and other analyses will be calculated in advance for each country or sector/field, respectively, and inserted into the software tool. This software tool will then provide a set of customized text modules and be able to automatically choose the appropriate formulation for the specific data-based content. In doing so, we will provide a unique software-based and automated approach which is able to generate continuous text as an input for the ID cards.

[36] Countries will be assigned, on the one hand, by the applicant address and, on the other hand, by the inventor address. This doubles the number of measures.

[37] Types of research performers are assigned according to the EEE-PPAT database provided by K.U. Leuven (Du Plessis et al. 2009; Magerman T. et al. 2009; Peeters B. et al. 2009).

[38] Triadic patents are defined as patent families with at least one family member at the EPO, the USPTO and the JPO (Grupp et al. 1996; OECD 2008). For the family definition, we use the INPADOC or the DOCDB definition as used in PATSTAT. We will <u>not</u> employ a nowcasting procedure.

[39] Transnational patents (Frietsch, Schmoch 2010) are defined as patent families with at least one family member at the EPO <u>or</u> WIPO (PCT). This is a more modern definition and does not require nowcasting. It is therefore more topical. For the family definition, we use the INPADOC or the DOCDB definition as used in PATSTAT.

[40] IP5 patents (Dernis et al. 2015) are defined as patent families with members in at least one of the IP5 offices (EPO, JPO, KIPO, SIPO and USPTO), and in any other office worldwide (anywhere in the world, not necessarily at another IP5 office).

[41] We define any patent as highly cited that belongs to the 10% most frequently cited patents in a 3 or 5 years citation window. This definition is similar to the Excellence Rate of the most highly cited publications (Waltman, van Eck 2013).

For the sectoral ID cards, additional indicators will be calculated, which will not be included in the database (input for the RIO) as such:

- Patent and R&D specialisation profiles,
- Shares of patents in worldwide patents,
- Number of patents per 1 million inhabitants/per 1 million employment,
- Shares of international co-patents,
- Shares of public-private co-patents.

References

Aschhoff, B.; Crass, D.; Cremers, K.; Grimpe, C.; Rammer, C.; Brandes, F.; Diaz-Lopez, F.; Klein Woolthuis, R.; Mayer, M.; Montalvo, C. (2010): European Competitiveness in Key Enabling Technologies. Final Report., Mannheim, Delft: ZEW/TNO.

Dernis, H.; Dosso, M.; Hervas, F.; Millot, V.; Squicciarini, M.; Verzzani, A. (2015): World Corporate Top R&D Investors: Innovation and IP bundles. A JRC and OECD common report , Luxembourg: Publications Office of the European Union.

Du Plessis, M.; Van Looy, B.; Song, X.; Magerman, T. (2009): Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation, EUROSTAT Working Paper and Studies Luxembourg.

Evenson, R.E.; Putnam, J. (1988): The Yale-Canada patent flow concordance No. Economic Growth Centre Working Paper, Yale.

Frietsch, R.; Neuhäusler, P.; Michels, C.; Bethke, N.; Rothengatter, O.; Eckl, V.; Kladroba, A.; Stenke, G. (2014): Identifikation der Technologieprofile von FuE-betreibenden Unternehmen anhand eines Matchings von FuE- und Patentdaten, Karlsruhe/Essen: Fraunhofer ISI / Stifterverband.

Frietsch, R.; Schmoch, U. (2010): Transnational Patents and International Markets. In: Scientometrics, 82, pp. 185-200.

Gkotsis, P. (2015): The capability of the EU R&D Scoreboard companies to develop Advanced Manufacturing Technologies - An assessment based on patent analysis, JRC Technical Report, Report EUR 27176 EN, European Commission, J.R.C.I.f.P.T.S.J.R.C. (ed.), Seville: IPTS.

Grupp, H.; Münt, G.; Schmoch, U. (1996): Assessing Different Types of Patent Data for Describing High-Technology Export Performance. In: OECD (ed.): Innovation, Patents and Technological Strategies: Proceedings of a Workshop held at OECD Headquarters in Paris, 8-9 December 1994. Paris: OECD, pp. 271-287.

Johnson, D.K.N. (2002): The OECD Technology Concordance (OTC): Patents by Industry of Manufacture and Sector of Use, STI Working Papers 2002/5 - DSTI(DOC(2002)5).

Lybbert, T.J.; Zolas, N.J. (2012): Getting Patents & Economic Data to Speak to Each Other: An 'Algorithmic Links with Probabilities' Approach for Joint Analyses of Patenting & Economic Activity, WIPO Economics & Statistics Series No. Working Paper No. 5, Geneva: WIPO.

Magerman T.; Grouwels J.; Song X.; Van Looy B. (2009): Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization, EUROSTAT Working Paper and Studies Luxembourg.

Neuhäusler, P.; Frietsch, R.; Bethke, N.; Mund, C. (2015): Assessing companies' capability to develop advanced manufacturing technologies in selected industrial sectors, Report to IPTS in the AMCAP project, Karlsruhe: Fraunhofer ISI.

OECD (ed.) (2008): Compendium of Patent Statistics 2008. Paris: OECD.

Peeters B.; Song X.; Callaert J.; Grouwels J.; Van Looy B. (2009): Harmonizing harmonized patentee names: an exploratory assessment of top patentees, EUROSTAT Working Paper and Studies Luxembourg.

Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organization (WIPO), Karlsruhe: Fraunhofer ISI.

Schmoch, U.; Gauch, S. (2004): Innovationsstandort Schweiz, Studie für das Bundesamt für Berufsbildung und Technologie, Karlsruhe: Fraunhofer-ISI.

Schmoch, U.; Laville, F.; Patel, P.; Frietsch, R. (2003): Linking Technology Areas to Industrial Sectors, Final Report to the European Commission, DG Research, Karlsruhe: Fraunhofer ISI.

Van de Velde, E.; Debergh, P.; Verbeek, A.; Rammer, C.; Cremers, K.; Schliessler, P.; Gehrke, B. (2013): Production and Trade in KETs-based Products: The EU Position in Global Value Chains and Specialization Patterns Within the EU, Brussels: European Commission, DG Enterprise.

van Looy, B.; Vereyen, C.; Schmoch, U. (2014): Patent Statistics: Concordance IPC V8 – NACE REV.2, Luxembourg: Eurostat.

Verspagen, B.; Morgastel, T.v.; Slabbers, M. (1994): MERIT concordance table: IPC – ISIC (rev. 2), MERIT Research Memorandum 2/94-004, Maastricht.

Waltman, L.; van Eck, N.J. (2013): Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. Centre for Science and Technology Studies. In: Scientometrics, 96 (3), pp. 699-713.

Annex III - Definition of SGCs

A shift in terms of application orientation is envisaged with Horizon 2020 (H2020) as part of the Innovation Strategy in European policy making. This implies a focus on pathways to technological solutions. Previous thematic priorities and research lines under FP6 and FP7 have been retained – the main change within H2020 is not the thematic priorities, but the type of projects funded. In an attempt to be less prescriptive, the Work Programmes (WP) under the Grand Challenges – which encompass previous thematic priorities to a large extent – try to indicate the areas expected to be targeted. This offers more flexibility and freedom than under FP7, but in terms of definitions and analyses, this openness is too broad or too unspecific to be operationalized and measured. There is no clear definition of what should be covered by the Societal Grand Challenges (SGCs) in terms of included or excluded scientific or technological fields. In consequence, there are no definitions or demarcations of the SGCs which could be directly used for their definitions in terms of patent classes.

As a starting point, we held a workshop in Brussels at the end of March 2015 with a number of experts on technology classifications, data treatment, and statistical analyses from a broad array of institutions including academia, the EPO, EUROSTAT, European Commission DG JRC as well as other policy makers from the Commission Services. Two main results emerged from the workshop. First, there is continuity in policy making that makes it possible to draw on previous analyses and findings, for example, in the context of FP7. A modular definition of technologies within SGCs allowed for a sufficient level of flexibility in defining the SGCs. Second the following procedure was established to elaborate a technology definition:

- Check the (current and maybe upcoming) work programme (WP) and legal basis of H2020 for funded technologies;
- Rely on existing classifications in the definition of the modular technological fields encompassed in the SGCs as far as possible
- Hold interviews and discussions for each SGC with representatives from the respective thematic directorates of DG RTD and for the security Grand Challenge, the responsible units from DG HOME and DG CONNECT to ensure an appropriate understanding of the SGCs;
- Use the information gathered to suggest technological definitions that are then re-discussed with the representatives.

In the course of these discussions with representatives, it became clear that it was also worthwhile to examine the newer WP to obtain a more general and longer-term perspective of each SGC rather than only the current WP. This was taken into account when fine-tuning and adapting the technology definitions. In addition, the structured interviews and discussions, during which first general technology lists were provided by the contractors (based on the available WPs of the SGC), led to a common understanding of the SGCs. The Commission representatives agreed to check the precise technology definitions suggested by the contractors after the interviews. All the representatives agreed that the view of SGCs should be broader and more general than in the current WP under H2020 and go beyond what is explicitly mentioned in the WPs. In addition, several SGCs address general issues some of which might not be directly funded under H2020, but through a Joint Undertaking (JU) like the Innovative Medicines Initiative (IMI) as well as private funding and research by companies. The latter is particularly important here as the core of the project for which this definition is developed deals with business R&D expenditures (BERD) and patent applications, about 90% of which originate in the private sector. **So this definition of the SGCs is in the broad sense of the challenges and not in the narrower sense of the funding programme.** We decided, in agreement with the representatives of the respective thematic directorates, that there could be some overlap between the definitions of the Societal Grand Challenges, i.e. the patents classified in each field do not have to be mutually exclusive.

The following Grand Challenges are addressed in the course of this working paper:

- Health, demographic change and well-being [HEALTH],
- Food security, sustainable agriculture, marine and maritime research, and the bio-economy [BIOECONOMY],
- Secure, clean and efficient energy [ENERGY],
- Smart, green and integrated transport [TRANSPORT],
- Climate action, resource efficiency and raw materials [CLIMATE],
- Secure societies – protecting the freedom and security of Europe and its citizens [SECURITY].

The Grand Challenge "Europe in a changing world – inclusive, innovative and reflective societies" cannot be included as it has few direct technological components, which can be searched in patent databases. Some other elements that fall within the remit of the Societal Grand Challenges as defined in Horizon 2020 could also not be covered in this definition. This is, for example, the case for Cultural Heritage which falls under the CLIMATE Grand Challenge. While Cultural Heritage has a technological dimension, it was deemed impossible to accurately capture this output. Some finely grained IPC classes could be assigned to this field, but most relevant patents would fall within IPC classes that are considerably broader. Including this field would therefore lead to a strong over or underestimation of technological output.

As far as possible, we have relied on existing patent classifications and assign them (totally or in parts) to the respective SGC rather than creating completely new ones.

For example, the OECD[42] provides definitions of selected fields, namely biotechnology, ICT, nano-technology and environmental technologies. Furthermore, several patent offices publish search strategies and technology definitions of particular fields in their annual reports, for example, electric and hybrid vehicle technologies, renewable energy technologies, or biotechnology.[43]

The EPO used to classify nano-technologies (Y01), climate mitigation technologies (Y02) and more recently also energy-saving technologies in smart grids (Y04) using their internal classification scheme, the ECLA (European Classification System) that supplemented the IPC (European Patent Office (EPO) 2013). The ECLA was then transferred to the CPC (Cooperative Patent Classification), which is jointly managed by the EPO and the USPTO. Y02 and Y04 technologies are still classified there[44], while Y01 was integrated into the IPC in 2010 and is now classified as B82. The SETIS[45] project as well as an EPO/UN[46] report on Climate Chance and Mitigation Technologies (CCMT) uses this classification so that it is well established in the field and accepted by stakeholders in the Commission Services as well as outside the Commission. In the context of these challenges – namely CLIMATE and ENERGY – we have also relied on the IPC Green Inventory by the WIPO[47].

In a project for DG-RTD called "Measurement and analysis of knowledge and R&D exploitation flows, assessed by patent and licensing data", INCENTIM at K.U. Leuven together with KITES (Geerts et al. 2011) at the University Bocconi developed a concordance between FP7 thematic priorities and IPC classes (Patent Indicators by Thematic Priority)[48]. They based their technological definitions on Schmoch (Schmoch

42 Please refer to http://www.oecd.org/innovation/inno/oecdworkonpatentstatistics.htm.
43 See for example the annual reports by the German Patent and Trademark Office at http://www.dpma.de/english/service/publications/annualreports/index.html.
44 In the call for this tender, this was mentioned as reference to the "IPC Green Inventory".
45 https://setis.ec.europa.eu/archive/project-mapping
46 http://www.epo.org/news-issues/press/releases/archive/2015/20151208.html
47 http://www.wipo.int/classifications/ipc/en/est/
48 Annelies Geerts, Gianluca Tarasconi, Francesca Innocenti, Xiaoyan Song, Julie Callaert, Maikel Pellens, Caro Vereyen, Cathy Lecocq, Stefano Breschi, Bart Van Looy (2011): Measurement and analysis of knowledge

2008) and assigned them to 15 thematic priorities. As they also kept a differentiation below the level of these 15 priorities, a modular definition of the thematic priorities was achieved, which were also used in this working paper to assign technologies to one of the SGCs.

These available classification schemes formed the basis for our definitions of the SGCs. The definitions were based on the version 2015.01 of the International Patent Classification (IPC). The IPC can be found, for example, on the website[49] of the World Intellectual Property Organization (WIPO). More recent versions of the IPC as well as concordance tables from previous to the current version can also be found there.

There were mainly four sources that were used for the definitions. These were the WIPO classification by Schmoch (2008), indicated as "WIPO" in the source column, while the green inventory[50] by the WIPO is explicitly mentioned as well. EPO's (2013) climate change and mitigation technologies are indicated as "EPO" and the biotechnology definitions rely on OECD so it is indicated as "OECD". Another source was the report provided by Geerts et al. (2011), which is then indicated as "Geerts et al.". In one case we also relied on some classes that were defined in the KETs Observatory (Van de Velde et al. 2013) mentioned as "KETs Observatory". Finally, a number of definitions were newly introduced or adapted. These latter ones are then mentioned as "own".

To stress it once again, we decided to allow for some overlap between the individual definitions of the Societal Grand Challenges, i.e. the patents classified in each field do not have to be mutually exclusive. This is why also some sub-fields show up in more than one SGC (e.g. Y02T in Climate and in Transport).

Definition of SGCs in terms of IPC classes

<u>1) Health</u>

FIELD TITLE	IPC CLASSES	SOURCE
E-health	G06Q50/22, G06Q50/24	own definition
Medical instruments	A61B, A61C, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G, H04R25, A61N1/39, A61B5	WIPO, 13 own definition
Pharmaceuticals	A61K (NOT A61K-008), A61P	WIPO, 16
Biotech definition	A01H1/00, A01H4/00, A61K38/00, A61K39/00, A61K48/00, C02F3/34, C07G11/00, C07G13/00, C07G015/00, C07K4/00, C07K14/00, C07K16/00, C07K17/00, C07K19/00, , G01N33/(53,54,55,57,68,74,76,78,88,92); (G01N27/327, C12M, C12N, C12P, C12Q, C12S) AND (A61K)	OECD

and R&D exploitation flows, assessed by patent and licensing data. Deliverable 1.6: Patent Indicators by Thematic Priority; Leuven, Milano.

49 http://web2.wipo.int/classifications/ipc/ipcpub/#refresh=page

50 http://www.wipo.int/classifications/ipc/en/est/

2) Bioeconomy

FIELD TITLE	IPC CLASSES	SOURCE
Agriculture/forestry	A01B, A01C, A01D, A01F, A01G, A01L, A01M, A61D, B02B, B29C, B29D, B29K, B29L, B99Z, C03B, C08J, C12L, , C13B5, C13B15, C13B25, C13B45	WIPO, 29
	A01P, C05B, C05C, C05D, C05F, C05G	parts of WIPO, 19
Pulp and paper	D21C, D, H	own definition based on WIPO
Machines (cartons, boxes, printing)	B31B, B31C, B31D, B31F, B41B, B41C, B41D, B41F, B41G, B41J, B41K, B41L, B41M, B41N, C14B, D01B, D01C, D01D, D01F, D01G, D01H, D02G, D02H, D02J, D03C, D03D, D03J, D04B, D04C, D04G, D04H, D05B, D05C, D06G, D06H, D06J, D06M, D06P, D06Q, D21B, D21C, D21F, D21G, D21J, D99	WIPO, 28, excl. texile machines
	A01G23/00, A01G25/00, E02D3/00	WIPO, green inventory [51]
Genetic engineering	A01H1/06, C12N15/00, C12N7/00	own definition
Landscape management	E02B3, E02D, E02F	own definition
Food	A01J, A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K, C13B10, C13B20, C13B25, C13B30, C13B35, C13B40, C13B50, C13B99	WIPO, 18
	A21B, A21C, A22B, A22C, A23N, A23P	Geerts et al.
(Future) Proteins	C07K	own definition
Biomass	C10L5/40, C10L5/42, C10L5/44, C10L5/46, C10L5/48, C10B53/02 A01C3/02, C02F11/04, C05F17/02, B01D53/84, F23G7/10	WIPO, green inventory own definition
Bio-materials	C08B, C08C, C08H, C09F, C11B, C11C, C13B, A01N, D21H, C08L1, C08L3, C08L5, C08L7, C09J101, C09J103, C09J105, C09J107, C09K17, A61K36/02, A61K36/03, A61K36/04, A61K36/05	own definition
Marine	A01H15	own definition
Biotech	A01H1/00, A01H4/00, A61K38/00, A61K39/00, A61K48/00, C02F3/34, C07G11/00, C07G13/00, C07G15/00, C07K4/00, C07K14/00, C07K16/00, C07K17/00, C07K19/00, G01N33/(53, 54, 55, 57, 68, 74, 76, 78, 88, 92);	OECD
	(G01N27/327, C12M, C12N, C12P, C12Q, C12S) AND NOT (A61K)	OECD, Eurostat
	C07C29, C07D475, C07K2, C08B3, C08B7, C08H1, C08L89, C09D11, C09D189, C09J189	additional codes from KETs Observatory
Animals/livestock management	A01K, A01M, A22B, A61D, A23N17	own definition
Household appliances (food-related)	F25D, A21B, A47J	own definition

[51] http://www.wipo.int/classifications/ipc/en/est/

3) Energy

FIELD TITLE	IPC CLASSES	SOURCE
CCMT	Y02, Y04	EPO

4) Transport:

FIELD TITLE	IPC CLASSES	SOURCE
Aeronautics	B64	WIPO, 32
Automobiles (cars and trucks)	B60	WIPO, 32
Trains	B61	WIPO, 32
Trailer and other wheelers	B62	WIPO, 32
Ships	B63	WIPO, 32
Logistics/handling:	B25J, B65, B66, B67	WIPO, 25
Safety	included in B60-B64	
Intelligent transport/navigation	G06Q10/08, G06Q50/28, G06Q50/30, G05D1/00, G06F17/00, G06F19/00, G01S, G08C, G08G, G01C	own definition
Infrastructure	E01B, E01C, E01D, E01F	WIPO, parts of 35
New power train	H01M	own definition
Bio fuels for transport	C07C67/00, C07C69/00; C10B53/02; C10G, C10L1/02, C10L1/14, C10L1/19, C10L3/00, C10L5/00, C10L5/40, C10L5/42, C10L5/44, C10L5/46, C10L5/48, C10L9/00; C11C3/10, C12M1/107, C12N1/13, C12N1/15, C12N1/21, C12N5/10, C12N9/24, C12P5/02, (C12P7/06 bis C12P7/14), C12P7/64, Y02E50, Y02E70/20; Y02E70/30	Geerts et al., based on green inventory
Characteristics of vehicles	F16H3, F16H48, H02K29/08, H02K49/10, F02B43, F02M21/02, F02M27/02, H02J7/00	WIPO, green inventory not covered by the classes used above
CCMTs in transportation	Y02T, B62C	EPO: Finding sustainable technologies in patents

5) Climate

FIELD TITLE	IPC CLASSES	SOURCE
Waste management and recycling	A43B1/12, A43B21/14, A61L11/00, A62D101/00; B01D 45-53/96; B03B 9/06; B03C 3/00; B09B; B09C; B22F 8/00, B29B 17/00; B62D 67/00; B63B 35/32; B63J 4/00, B65F; B65G 5/00, C01B 31/20; C02F, (C04B7/24 bis C04B7/30), (C04B18/04 bis C04B18/10), C05F, C08J11/00, (C09K3/22 bis C09K 3/32), C09K11/01; C10B21/18; C10G1/10, C10L 5/46, C10L5/48, C10L10/02, C10L10/06, C11B 11/00, (C11B13 bis C11B13/04), C14C3/32, C21B3/04; C21C5/38; C22B7/00-7/04 C22B19/30, C22B25/06; C25C1/00; D01F13/00; D01G11/00; D21B1/08, D21B1/32; D21C5/02; E02B15/04, E02B15/08; E03C1/12; E03F, E21B41/00, E21B43/16; E21F17/16; F01N 3/00-3/38, F01N9/00; F02B75/10; F23B80/02; F23C9/00; F23J; F25J3/02; G08B21/12; G21F9/00; H01J9/50, H01J9/52; H01M6/52, H01M10/54; Y02E20/12, Y02C, B23D25/14, D21C11/00 B03B, B07B, B29B	Geerts et al., based on WIPO, green inventory own definition
Water and wastewater	B63B35/32, B63J4, C02F, C05F7, C09K3/32, E02B15/04, E03C1/12, E03F, G21C13/10, E03B	WIPO, green inventory own definition
Air	B01D45/00, B01D46/00, B01D47/00, B01D49/00, B01D50/00, B01D51/00, B01D53/00, B03C3/00, C10K, C10L10/02, C21B7/22, C21C5/38, C21C5/40, F01N3, F01N9, F02M27, F23C9/06	own definition
Air quality management	B01D45, B01D46, B01D47, B01D49, B01D50, B01D51, B01D53, B03C3/00, C09K3/22, C10B21/18, C10L10/02, C10L10/06, C21B7/22, C21C5/38, F01N3/00, F01N9/00, F02B75/10, F23B80/02, F23C9/00, F23G7/06, F23J7/00, F23J15/00, F27B1/18, F27B15/12, G08B21/12	WIPO, green inventory
Soil	C09K17/00, E02D3/00, C05F, B09C	WIPO, green inventory own definition
Noise	B25D17/11, B25D17/12, B60R13/08, B64C1/40, B64F1/26, E01B19, E01C1, E01F8, E04B1/74, (E04B1/80, 82, 84, 86, 88, 90), E04F15/20, E06B5/20, F01B31/16, F01N1, F01N13/02, F01N13/04, F02B77/13, F02C7/045, F02C7/24, F02K1/34, F02K1/44, F02K1/46, F02M35/12, F02M35/14, F16L55/02, F41A21/30, G10K11/16	own definition
Forests, flora, fauna	A01G23, A01H	own definition
Other	E03B3, B64G1/10	own definition
Bio-materials:	C08B, C08C, C08H, C09F, C11B, C11C, C13B C08L1, C08L3, C08L5, C08L7, (C09J101, 103, 105, 107)	own definition

6) Security:

FIELD TITLE	IPC CLASSES	SOURCE
Detection	G01B9/02, G01J3/45, G01J5/02, G01N29/00, G01N33/22, G01N33/569, G01N33/94, G01R27/00, G01V, G03B42/02, G21K1/00	Geerts et al. own definition
Forensics	A61B5/117, G01N33, G06M11/02	own definition
Monitoring/Navigation	B63G8/39, G01S13/00, G01S5/00, G08B13/00, G08B15/00, G08B17/00, G08B19/00, G08B21/00, G08B23/00, G08B25/00, G08B26/00, G08B27/00, G08B29/00, G08B31/00, B60R25/00	own definition
Access control	A61B 5/117, E05B39/00, E05B45/00, E05B75/00, G06F21/00, G06F7/04, G06K5/00, G06K7/00, G06K9/00, G07C9/00, H04L9/00, H04W12/00	own definition
Protection	A62, F42D5, E04H9, H05K9, B60R21/12	own definition
Protective clothing	A41B9/12, A41D13/00, B63C9/08, B63C11/00, B64G6/00, G21F3/00	own definition
Equipment	A62B, E21F, A63B29/02, B63C9/00, E04H9/00,	from HTS
Catastrophe fighting	A62C, A62D, B09C, G01J5/00, G01T1/00	own definition
Public communication	H04K, H04L 9/00	own definition
Critical infrastructure	B61L23/04, B61L29/02, C04B111/20, E01F13/00, E01F15/00, E02B3/04, E04H9/00, E06B5/10, F41H5/00, G08B13/00, G08B15/00, G08B17/00, G08B19/00, G08B21/00, G08B23/00, G08B25/00, G08B26/00, G08B27/00, G08B29/00, G08B31/00, G21F7/00, H02J9/00, H05C	own definition
Digital security	G06F12/14, G06F21/00, H04L9/00, G07F7/08, G07F7/10, G07F7/12, G06F13/362, G06Q20/40, G06K9/00, H05K 9/00, G09C	own definition

References

European Patent Office (EPO) (2013): *Finding sustainable technologies in patents*: European Patent Office.

Geerts, A.; Tarasconi, G.; Innocenti, F.; Song, X.; Callaert, J.; Pellens, M.; Vereyen, C.; Lecocq, C.; Breschi, S. and van Looy, B. (2011): *Measurement and analysis of knowledge and R&D exploitation flows, assessed by patent and licensing data. Deliverable 1.6: Patent Indicators by Thematic Priority*. Leuven and Milano.

Grupp, H.; Münt, G. and Schmoch, U. (Eds.) (1996): *Assessing Different Types of Patent Data for Describing High-Technology Export Performance*, OECD.

Schmoch, U. (2008): *Concept of a Technology Classification for Country Comparisons: Final Report to the World Intellectual Property Organisation (WIPO)*. Geneva: World Intellectual Property Organization.

Schmoch, U.; Laville, F.; Patel, P. and Frietsch, R. (2003): *Linking Technology Areas to Industrial Sectors, Final Report to the European Commission, DG Research*. Karlsruhe: Fraunhofer ISI.

Van de Velde, E.; Debergh, P.; Verbeek, A.; Rammer, C.; Cremers, K.; Schliessler, P. and Gehrke, B. (2013): *Production and Trade in KETs-based Products: The EU Position in Global Value Chains and Specialization Patterns Within the EU*. Brussels: European Commission, DG Enterprise.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: http://europea.eu/contact

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: http://europa.eu/contact

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU publications

You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact).

KJ-07-17-111-EN-N

JRC Mission

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

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/39818

ISBN 978-92-79-76847-7