The Development of Digital Sustainability Technologies by Top R&D Investors

Jindra, Björn ; Leusin, Matheus

Document Version
Final published version

DOI:
10.2760/150239

Publication date:
2022

Citation for published version (APA):
Jindra, B., & Leusin, M. (2022). *The Development of Digital Sustainability Technologies by Top R&D Investors*. Publications Office of the European Union. https://doi.org/10.2760/150239

European Commission

THE DEVELOPMENT OF DIGITAL SUSTAINABILITY TECHNOLOGIES BY TOP R&D INVESTORS

Jindra, B.

Leusin, M.

2022

Joint Research Centre

Contact information
Name: Francesco Rentocchini
Address: Av. Inca Garcilaso 3, 41092 Seville, Spain
Email: francesco.rentocchini@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC130480

PDF ISBN 978-92-76-56422-5 doi:10.2760/150239 KJ-07-22-942-EN-N

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Jindra, Björn., Leusin, Matheus., The Development of digital sustainability technologies by top R&D investors, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/150239, JRC130480.

Contents

Abstract

This report focuses on inventions related to digital sustainability technologies, which develop at the interface of the green and digital transition. For the observation period 2001 to 2018, the report identifies four main clusters of related technologies in the creation of digital sustainability inventions: (1) ‘Energy generation and data-related technologies’, (2) ‘Technologies related to the capture, storage, sequestration or disposal of greenhouse gases’, (3) ‘Technologies related to the processing of goods and domestic applications’, and (4) ‘Technologies related to transportation’. Although, Cluster (1) unites the highest number of specialised technologies, the growth of digital sustainability technologies in Cluster (3) is most dynamic. For the period from 2016 to 2018, the report documents that digital sustainability inventions concentrate on merely 403 out of 2000 R&D scoreboard companies, which account for 64.8% of all digital sustainability patents in the observation period. Digital sustainability patents represent only 1.6% of all patent filings and 14.1% of all sustainable patents by these firms. By this measure, the so called ‘twin transition’, based on the integration of green and digital technologies, is still in an early phase. The majority of digital sustainability patents concentrate amongst ‘ICT producers’, which refers specifically to R&D scoreboard companies in ‘Technology Hardware & Equipment’ and ‘Electronic & Electrical Equipment’ sectors. Despite the fact that the European R&D scoreboard companies lead in the number of specialisations in technologies relevant to the deployment of digital sustainability inventions, they lag behind in terms of output of digital sustainability inventions and contribute primarily to Cluster (3) and (4). European R&D scoreboard companies might lack important specialisations, for example, in alternative energy sources, like solar and nuclear power as well as data and ICT-related technologies and electrical vehicles.

Executive summary

Arguably, the next decades will be characterised by the twin green and digital transitions. Major aspects of these transitions need to be subject to impact assessment for review by policy-makers and relevant stakeholders. This report provides novel insights about the development of digital sustainability technologies, which materialize at the interface of the green and digital transitions. We develop a method to identify inventions associated with digital sustainability technologies, investigate their nature and evolution, and explore the role of the world's top corporate R&D investors in the development of digital sustainability technologies.

Focusing on digital sustainability inventions in the period between 2001 and 2018,

- We suggest that the **core of digital sustainability technologies** is composed of '*Climate change mitigation technologies related to transportation*', '*Climate change mitigation technologies in ICTs*', '*Climate change mitigation technologies related to buildings*', '*Reduction of GHG emissions, related to energy generation, transmission or distribution*', as well as '*Climate change mitigation technologies in the production or processing of goods*'.

- We find the **majority of digital sustainability inventions are filed in** *'Climate change mitigation technologies in the production or processing of goods'.* However, digital sustainability inventions related to *'Climate change mitigation technologies in ICT'* have their protection proportionally most often extended to other patent offices, and therefore, seem to diffuse most in terms of geography.

Furthermore, we identify **four main clusters of related technologies** in the creation of digital sustainability inventions, which have been emerging with distinct dynamics in the period 2001 to 2018:

- Cluster 1: *'Energy generation and data-related technologies'*, Cluster 2: '*Technologies related to the capture, storage, sequestration or disposal of GHGs'*, Cluster 3: *'Technologies related to the processing of goods and domestic applications'*, and Cluster 4: *'Technologies related to transportation'*.
- **Cluster 1 unites the highest number of specialized technologies.** Cluster 3 and 4 have been increasing in the number of specialised technologies, whereas Cluster 2 has been losing specialised technologies over time. **The growth trajectory of digital sustainability technologies in Cluster 3 seems the most dynamic**.

For the most recent observation period from 2016 to 2018,

- We find that while **R&D Scoreboard companies** have been responsible for filing 60.6% of priority patents, this proportion reaches **64.8% for digital sustainability technologies**. This hints at the importance of R&D Scoreboard companies in the era of the twin transition.
- However, **only 1.55% of all priority patents by R&D Scoreboard companies** and 14.11% of all their sustainable priority patents **are digital sustainability inventions**. By this measure, this implies that the twin transition is still in an early phase and the integration of green and digital technologies is still very limited.

In terms of geography, only **25 out of 40 countries host R&D Scoreboard companies with at least one digital sustainability invention**.

- The **Top 5 countries** are the **US, Japan, China, South Korea, and Germany**. Amongst the Top 15 countries, we also find other European locations such as France, Netherlands, Sweden, the UK, and Switzerland. In the case of Ireland and Sweden, their proportional contribution to digital sustainability inventions exceeds their respective contribution to all inventions.

Only 20% of the 2,000 R&D Scoreboard companies have at least one digital sustainability invention.

- The Top 50 and **Top 10 of the R&D scoreboard companies account** for 66.2% and **32.3% respectively of digital sustainability inventions**.

- Amongst the Top-10, we find Qualcomm, Intel, and General Electric (all US), Samsung Electronics (KR), Fanuc (JP), Siemens (DE), Ford Motor (US), Hitachi (Japan), Zte (CN), and Huawei (CN). **Amongst, the Top-50 R&D we find only four companies from EU countries**: Siemens (DE), BMW (DE), Schneider (FR), and Johnson Controls (IE).

In terms of economic sectors, **'ICT producers', 'Industrials', and 'Automobile and other transports' produce most digital sustainability inventions**.

- The majority of digital sustainability inventions are concentrated amongst 'ICT producers', which refers specifically to **'Technology Hardware & Equipment'** and '**Electronic & Electrical Equipment' sectors**.
- The **core of digital sustainability inventions originates from digital-related companies embracing sustainable technologies**, rather than the other way around. This finding points to the importance of digital industries to advance the twin transition.

Considering **technological specialisation advantages based on all inventions by R&D Scoreboard companies** for the period 2016-2018,

- We find that R&D Scoreboard companies from the US and China, are specialised in '*Climate change mitigation technologies in ICTs*' as well as in related technologies such as *'Electrical digital data processing*', '*Data processing systems or methods*', '*Transmission of digital information*', and '*Wireless communication'* (Cluster 1).
- **European** and Japanese **R&D Scoreboard companies share specialisations in *'Technologies related to the processing of goods and domestic applications'* as well as related technologies** such as '*Working metallic powder'*, *'Additive manufacturing'*, '*Healthcare informatics'*, and *'Superheating of steam'* (Cluster 3).
- In turn, **R&D Scoreboard companies from the US, Japan, and the EU lead in specialisation advantages linked to *'Technologies related to transportation',* sharing many specialisations in related technologies** (Cluster 4).

Investigated the technological specialisation of R&D scoreboard companies in the main clusters,

- We find that although R&D scoreboard companies own particularly large patent portfolios, they concentrate their specialisations in specific clusters of digital sustainability technologies.
- **The majority of the 66 unique R&D scoreboard companies with the highest technological specialisation originate from Japan, followed by the US, China, Germany, and Taiwan**.
- At the EU27 level, there are only 11 European scoreboard companies. The European R&D scoreboard companies contribute primarily in 'Technologies related to the processing of goods and domestic applications' (cluster 3) and '*Technologies related to transportation'* (Cluster 4).

Despite the fact that the **European R&D scoreboard companies lead in number of specialisations in technologies relevant to the deployment of digital sustainability inventions, they lag behind R&D scoreboard companies from the US and Japan in terms of output**. This could be linked to missing relevant technological specialisations.

- For '*Energy generation and data-related technologies*' (Cluster 1), **European R&D companies miss specialisations in other alternative sources of energy, like solar and nuclear power-related technologies**, in 'Transmission systems for measured values, control or similar signals' as well as in 'Reduction of GHG emissions related to energy generation, transmission or distribution'. In particular, they **lack most data and ICT-related specialisations,** such as 'Electric digital data processing', 'Data processing systems or methods', 'Wireless communication networks', and 'CCMTs in ICT'.
- In '*Technologies related to transportation'* (Cluster 4), European R&D scoreboard companies possess most of the relevant specialisations apart from **specialisations related to electrical vehicles**, which is the second largest and fastest-growing technology of this cluster.

A **European policy strategy could focus**, for example, upon

- Leveraging the further development of technological specialisations linked to alternative sources of electric power generation and to the transmission/distribution and management of electric power systems to improve the performance in **digital sustainability inventions in the area of energy generation and data-related technologies**. Through related diversification, this would allow for the development of missing specialisations in technologies linked to data processing and ICTs.
- Further **development of specialisations related to digital sustainability inventions in transportation technologies** could favour the emergence of new specialisations required for the deployment of **electric vehicle-related inventions**.

Part I: Main Report

1. INTRODUCTION

The next decades will be deeply characterised by the twin green and digital transitions. Major aspects of these transitions need to be subject to impact assessment for review by policy-makers and relevant stakeholders. This report provides novel insights on digital sustainability technologies, which develop at the interface of the green and digital transition. The notion of 'twin transition' refers to the potential of digital technologies to enable a more sustainable future through increasing energy and resource efficiency (Amoroso et al., 2021). Arguably, this transition depends on the deployment of digital sustainability technologies, which combine digital and sustainability-related components.

Conceptually, digital sustainability technologies can be defined as the technologies that allow us 'to create, use, and regulate digital resources in order to maximize their value for our society today and in the future' (Stuermer, 2014). George et al. (2021) argue that digital sustainability technologies 'seek to advance the sustainable development goals through creative deployment of technologies that create, use, transmit, or source electronic data'. Thus, digital sustainability technologies combine a 'digital' component with a 'sustainable' one. Each component on its own is subject to extant research, with a variety of search strategies proposed to identify them individually (see for example Inaba and Squicciarini 2017; OECD 2019; Sadowski et al. 2016; EPO 2020; León et al. 2018; OECD 2016). However, so far we lack approaches to identify digital sustainability technologies. In this context, this report develops a method to identify patents associated with digital sustainability technologies to investigate their nature and evolution over time.

We propose an identification strategy based upon six search modules, which combine specialists' opinions, keywords, and classification-based approaches. We use the Y section of the Cooperative Patent Classification (CPC) (USPTO, 2021) as our main reference. The scheme, launched in 2013 by a joint effort between the EPO and USPTO, combines algorithm-based identification with specialists' opinions (Angelucci et al., 2018). There are alternatives to the 'Y' scheme that also aim at identifying green technologies such as the OECD ENV-TECH classification (see Hascic and Migotto, 2015; OECD, 2016). Since the OECD ENV-TECH classification uses exclusively codes to identify green technologies, the Y scheme offers an additional quality filter through the use of specialists' opinion, which lowers the occurrence of false positives. In addition, the Y-selected dataset offers better coverage of digital technologies linked to the technological fields of 'Digital communication', 'Computer technology', and 'Telecommunications' (for a comparison see Section 2.2 Part II). The 'Y' scheme, therefore, identifies very accurately sustainable technologies as 'technologies or applications for mitigation or adaptation against climate change' (Y02 code) as well as 'promising' digital technologies as 'ICT technologies having an impact on other technology areas' (Y04 code). We consider the 'Y' classification scheme as preferable over the alternative OECD ENV-TECH classification (OECD, 2016) when one is particularly interested in identifying technologies that combine both sustainability and digital components. The trade-off is having a lower patent count (see for detailed discussion, Part II, Section 2.2).

Yet, the Y scheme on its own is not enough to identify technologies that are both sustainable and digital. Although it offers a code to identify digital technologies, it introduces a bias toward digital technologies related to electric power, since the Y04 code refers exclusively to 'Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grids' (see for a detailed discussion Part II, Section 2.1). Therefore, we extend our search by using keywords related to digital technologies and applying International Patent Classification (IPC) codes identified as typical for digital technologies to collect patents classified under the Y02 sustainability tag (for more details see Part II, Section 1). Widening the scope of the search, we improve especially the coverage of 'Climate change mitigation technologies in the production or processing of goods' as well as 'Climate change mitigation technologies ICTs aiming at the reduction of their own energy use'. Alternative search strategies (see for example, Amoroso et al., 2021) rely exclusively on an ICT classification (Inaba and Squicciarini, 2017) to capture the digital component in the identification of digital sustainability patents. In comparison, our strategy aims to include ICT-technologies without being exclusively restricted to a classification. Therefore, we combine keyword-based strategies with classification-based strategies to capture digital components in digital sustainability inventions more broadly.

We apply our search strategy to PATSTAT 2019a and identify 319,243 patents associated with digital sustainability technologies. The reported accuracy is above 95.5% for all search modules considering their efficacy in capturing patents that are both digital and sustainable (for more details see Part II, Section 2.1). We use this first dataset in a technology space approach to analyse the nature and evolution of digital sustainability inventions from a technological perspective for the period from 2001 to 2018. First, we use a technology space approach to identify nodes (Y02 subclasses) with a degree of high connectivity to other nodes, which implies they are most frequently used to create digital sustainability inventions. In the subsequent analysis, we describe patent sections surrounding these nodes and investigate, whether we observe particular commercialisation or diffusion patterns by distinguishing priorities and non-priorities for Y02 subclasses. Second, we apply the Revealed Technological Advantage (RTA) index to identify technologies that are relatively often used to create particular digital sustainability inventions during the observation period (see Part II, Section 1.3). We assume that technologies within a specific cluster are more often combined with each other to create particular digital sustainable inventions than technologies outside the respective cluster. On this basis, we identify four main clusters of related technologies. In the subsequent analysis, we look at the evolution of these clusters over three time intervals from 2001 to 2018, providing an overview of constituting specialised technologies as well as examples of inventions for each of the clusters.

In the next section, we investigate the contribution by the world's top corporate R&D investors towards the development of digital sustainability inventions. We apply now the proposed patent identification strategy to PATSTAT 2021b. This update not only increases the number of digital sustainability patents identified to 494,255, but also allows improved coverage of more recent years. We identify priority patents using the IP5 strategy for the period from 2016 to 2018 (see Part II, Section 3). Using the IP5 strategy excludes patents filled in just one patent office and generates a dataset of related inventions with a 'comparable' technological and economic value (Dernis et al., 2015). Given that R&D Scoreboard companies have large, complex, and especially changing ownership structures due to exits, M&As, etc., we need to restrict the firm level analysis to a relatively short period of 3 years. We link patents to R&D Scoreboard companies using the JRC-OECD COR&DIP© v.3 dataset (Hernández et al., 2020; Amoroso et al., 2021).

In the subsequent analysis, we first investigate the output of digital sustainability inventions by R&D Scoreboard companies. Thereby, we consider their global contribution in comparison to other actors and offer a breakdown of the output of digital sustainability priorities of R&D Scoreboard companies by country of origin, sector (Scoreboard, ICB, and NACE classification), as well as firm level. Second, we investigate the specialisation of R&D Scoreboard companies in technologies relevant to the deployment of digital sustainability inventions as previously identified. The premise is that, due to knowledge commonalities (Breschi et al., 2003), having a specialisation in relevant technologies indicates that R&D Scoreboard companies command the required knowledge to deploy digital sustainability inventions, even if the corresponding output might not yet materialise. Therefore, the subsequent analysis considers all priority patents registered by R&D Scoreboard companies for the observation period (2016-2018). In the analysis, we consider the technological specialisations of all R&D Scoreboard companies jointly and the specialisation of individual firms. Finally, we focus on the specialisation pattern observed for the European R&D Scoreboard companies.

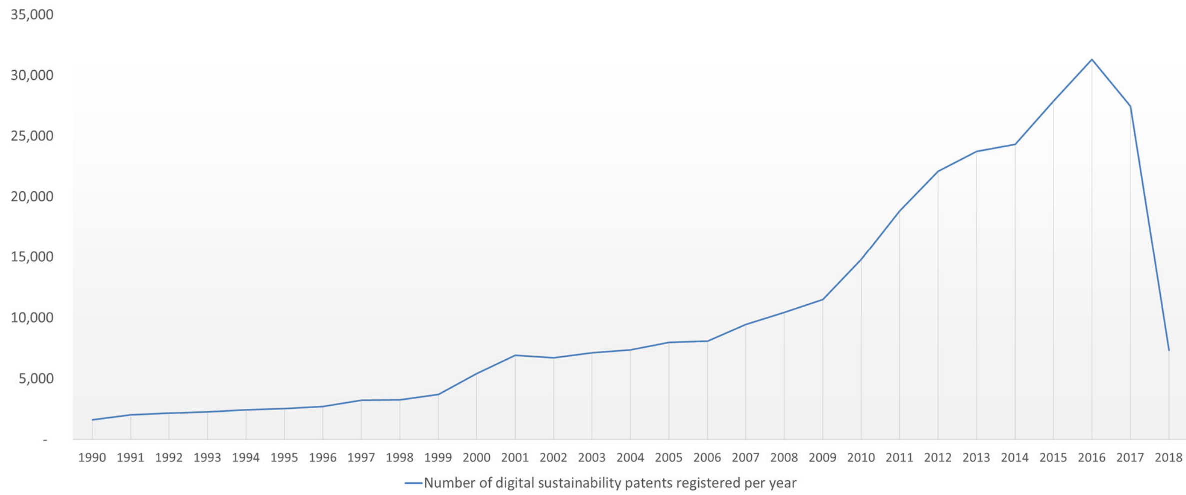
After this introduction, Part I of this report provides an analysis of digital sustainability technologies (Section 2). Based on the technology space approach, we explore the composition and evolution of digital sustainability technologies (Section 2.1) and identify the main clusters of relevant digital sustainability technologies (Section 2.2). In Section 3, we focus on R&D Scoreboard companies, analysing their output of digital sustainability inventions as well as their existing and missing specialisations in technologies relevant to the deployment of digital sustainability inventions. Section 4 summarises the main findings. Section 5 discusses policy implications. Part II of this report offers a Methodological Note. It provides details on the methods applied in the identification strategy (Section 1), the validity of the search results (Section 2), information on the matching of the patent data with R&D Scoreboard companies (Section 3), as well as information on access to data and code (Section 4).

2 DIGITAL SUSTAINABILITY TECHNOLOGIES

2.1 Composition and evolution of digital sustainability technologies

The analysis presented in this report combines six search modules in a strategy to identify patents related to digital sustainability technologies (see Section 1 in Part II). Applying the identification strategy to PATSTAT 2019a generates a total recall of 319,243 patents associated with digital sustainability technologies. Out of these only 4.6% were registered before 1990. Figure 1 presents the annual number of digital sustainability priority filings for the period from 1990 to 2018.

Figure 1: Number of digital sustainability patents registered between 1990 and 2018.



*Source***:** PATSTAT 2019.

We observe an increase in the rate of new filings at the beginning of the 2000s. Thus, digital sustainability inventions mainly developed in the last two decades. The rate of new filings increases especially after 2009. In fact, 65.5% of all identified digital sustainability priority filings happened between 2009 and 2018. The decline in the last two years (2017 and 2018) is explained by a delay in the registration from all patent offices considered in PATSTAT 2019a.

To understand the composition and evolution of digital sustainability technologies, we apply a technology space perspective. This approach allows a data-driven identification of relevant technologies without any ex-ante assumption about how they relate to each other. This method shows how distinct technologies are combined together in a same invention through the links between nodes. The combinations are used to calculate the degree of connectivity of each node. Similar technologies are placed closer to each other in a network visualisation; conversely, the greater the distance between two technologies, the less similar they are. In the calculation of the technology space we follow Breschi et al. (2003), which propose a relatedness measure that is normalised using the cosine index. Applying the technology space perspective allows visualising how distinct technologies, used to create digital sustainability patents, relate to each other. In our application, we rely on patent classification codes as a proxy for technologies. However, we need to acknowledge that classification codes not necessarily have clear distinct boundaries between each other, and that their use is up to the subjective interpretation of the examination specialists at respective patent offices.

The calculation of the technology space uses 168,353 digital sustainability inventions classified as priority patents1 (2001 – 2018). Using priority patents avoids that the same invention is counted more than once, which would potentially bias the results towards patents that are registered in more than one patent office. We use 4-digit level

[1] A priority filing (or priority patent) is the first patent application filed to protect an invention. In case, the same patent is registered in other patent offices, the subsequent registrations are called non-priorities, constituting a patent family linked through the priority filing. From the identified 319,243 digital sustainability patents, 194,440 are priorities. 168,353 from these were registered in the 2001-2018 period considered.

CPC codes assigned to patents to proxy the technologies due to their availability in the 'Y' classification scheme[2]. As outlined above, the CPC classification is especially useful for the purpose of the technology space visualisation considered here, since it allows identifying separately how green technologies integrate with other technologies. The green technologies considered in the Y02 code refer to technologies or applications for climate change adaptation and mitigation. In the visualization of the technology space (see Figure 2), we highlight the eight subclasses related to the Y02 classification (see Table 1).

Table 1: Y classification for subclasses linked to code Y02

Code	Sub-classification	Description
Y02	**Technologies or applications for mitigation or adaptation against climate change.**	
	Y02A	Technologies for adaptation to climate change.
	Y02B	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications.
	Y02C	Capture, storage, sequestration or disposal of greenhouse gases [GHG].
	Y02D	Climate change mitigation technologies in information and communication technologies [ICT], i.e. information and communication technologies aiming at the reduction of their own energy use.
	Y02E	Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution.
	Y02P	Climate change mitigation technologies in the production or processing of goods.
	Y02T	Climate change mitigation technologies related to transportation.
	Y02W	Climate change mitigation technologies related to wastewater treatment or waste management.

Source: USPTO (2021).

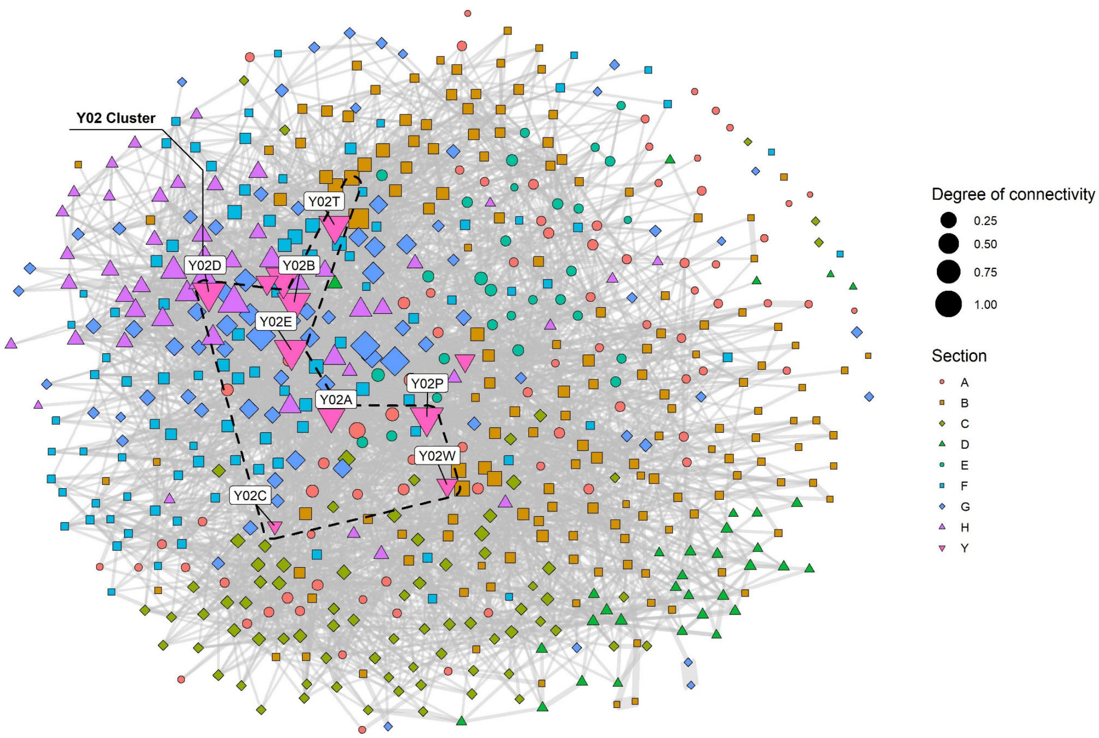
We find the nodes related to the Y class in a central position of the technology space. They also show a high degree of connectivity to other nodes. This implies that they are at the core of digital sustainability technologies and frequently used to create inventions related to these technologies. In particular, nodes linked to 'Climate change mitigation technologies related to transportation' (Y02T), 'Climate change mitigation technologies in ICT' (Y02D), 'Climate change mitigation technologies related to buildings' (Y02B), 'Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution' (Y02E), and 'Climate change mitigation technologies in the production or processing of goods' (Y02P) are most frequently used to create digital sustainability inventions.

The technology space also shows clusters of CPC sections surrounding some of the Y02 codes. Hence, technologies that compose these sections are very similar to the respective Y02 subclasses that they surround. In particular, the subclass 'Climate change mitigation technologies in ICT' (Y02D) is surrounded by codes from Section H (Electricity); subclasses 'Climate change mitigation technologies related to transportation' (Y02T) and 'Climate change mitigation technologies related to wastewater treatment or waste management' (Y02W) are closer to codes from Section B (Performing Operations; Transporting), and subclass 'Capture, storage, sequestration or disposal of GHG' (Y02C) is closer to codes from Section C (Chemistry; Metallurgy). Moreover, we find many codes related to sections G (Physics) and F (Mechanical Engineering; Lighting; Heating; Weapons; Blasting) in the centre and vicinity of the Y02 cluster, whereas codes related to section A (Human necessities) are more dispersed throughout the network. Codes related to section D (Textiles; Paper) and E (Fixed constructions) are distanced from the Y02 technologies, mostly in the bottom-right and upper-right parts of the network, respectively. Thus, CPC codes related to A (Human necessities), section D (Textiles; Paper) and E (Fixed constructions), have been playing a minor role in the development of digital sustainability inventions so far.

Figure 2:

[2] Complete Y scheme available at https://www.uspto.gov/web/patents/classification/cpc/html/cpc-Y.html.

Figure 2: Technology Space of Digital Sustainability Technologies for the considered period (2001-2018).



Source: PATSTAT 2019a.

Note: In the above visualisation, the shape of nodes represent the CPC section of each node. The distribution of nodes visualises their relatedness: Nodes close to each other are more similar, than nodes that are more distant to each other. The links between nodes indicate how often they co-occur in the same patent. The size of the nodes represents how often they are combined with other technologies (i.e., their degree of connectivity): The more often a node is combined with other technologies, the larger is the node. Nodes related to the Y02 codes are highlighted with the label of the respective subclasses.

Next, we exploit information on the frequency of priority and non-priority digital sustainability patents (2001 – 2018) across all subclasses linked to climate change mitigation and adaptation technologies (i.e., through the Y02 code).

Table 2: Number of occurrences and share of Y02 subclass in all digital sustainability patents (2001-2018).

	Priorities		Non-priorities			
Y02 Code	Number of occurrences	Share	Number of occurances	Share	Priorities minus Non-priorities	Ratio Non-priorities to priorities
Y02A	12,222	7%	3,939	4%	3%	0.32
Y02B	30,973	18%	16,754	15%	3%	0.54
Y02C	90	0%	93	0%	0%	1.03
Y02D	32,829	19%	39,966	36%	**-17%**	1.22
Y02E	31,885	18%	15,437	14%	4%	0.48
Y02P	49,391	28%	21,525	19%	**9%**	0.44
Y02T	16,468	9%	13,804	12%	3%	0.84
Y02W	2,991	2%	1,004	1%	1%	0.34

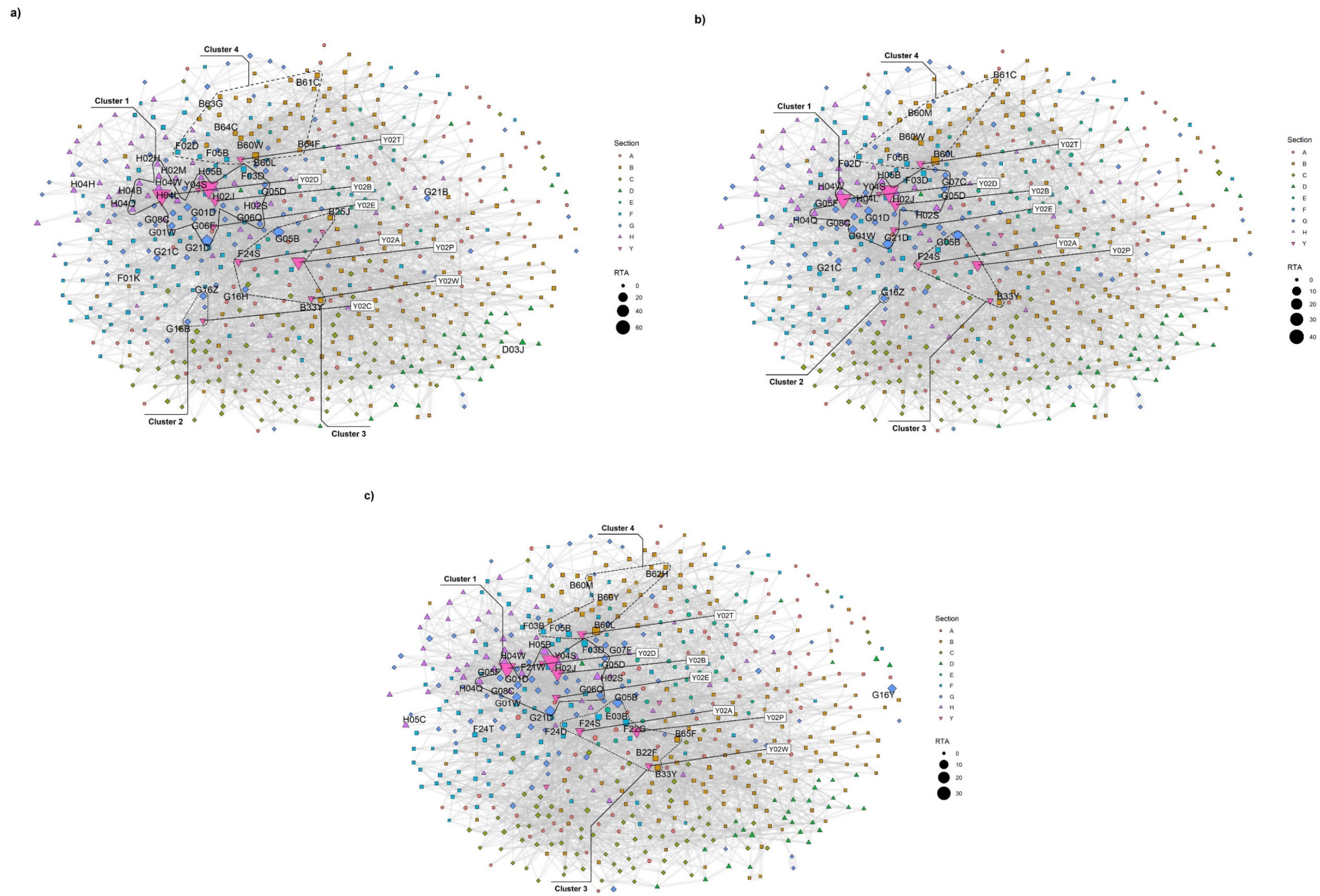
Source: PATSTAT 2019a.

In terms of the occurrences, we find the majority of digital sustainability priorities in ‘Climate change mitigation technologies in the production or processing of goods’ (Y02P), ‘Climate change mitigation technologies in ICT’ (Y02D), ‘Reduction of GHG emissions, related to energy generation, transmission or distribution’ (Y02E), and ‘Climate change mitigation technologies related to buildings’ (Y02B). Comparing priorities to non-priorities, we can see that the commercialisation or diffusion patterns differs across subclasses. For example, priorities linked ‘Climate change mitigation technologies in ICT’ (Y02D) have their protection proportionally more often extended to other patent offices in comparison to all other subclasses. Although, patents in ‘Climate change mitigation technologies in the production or processing of goods’ (Y02P) occur most frequently, their protection is less often extended to other patent offices. Thus, priorities linked ‘Climate change mitigation technologies in ICT’ are diffusing internationally, whereas digital sustainability inventions in ‘Climate change mitigation technologies in the production or processing of goods’ as well as in ‘Reduction of GHG emissions, related to energy generation, transmission or distribution’ are more restricted to the jurisdiction of the initial (domestic) patent office.

2.2 Main clusters of relevant digital sustainability technologies

We use the technology space approach to show how distinct specialisations related to digital sustainability patents emerged and evolved. In particular, we use the Revealed Technology Advantage (RTA) Index as a specialisation measure to highlight technologies that are relatively frequently used to create digital sustainability inventions (see Box 1). This allows conclusions as to which technologies are more important to the development of digital sustainability inventions and how this importance changed over time. Thereby we can identify ‘clusters’ of relevant technologies based on the position of these specialisations in the technology space. Figure 3 presents these specialisations for three distinct time intervals: 2001-2006, 2007-2012, 2013-2018. For better visualization, we highlight only codes with a specialisation (i.e., RTA >= 1), with the Y02 codes being separately marked, if they have a specialisation. The size of the nodes now represents the value of the specialisations linked to them.

Figure 3: Specialisations of Digital Sustainability Technologies for the intervals 2001-2006 (a), 2007-2012 (b), and 2013-2018 (c)



Source: PATSTAT 2019a (for defining the technology space) and PATSTAT 2021b (for calculating specialisations).

Note: The technology space is the same as presented in Figure 2, which was calculated based on PATSTAT 2019a. The 4-digit CPC code must have been used at least once in a digital sustainability patent (2001-2018) to be shown in the technology space. Conversely to the previous Figure 2, we now use CPC codes labels to highlight the presence of a specialisation (instead of the Y02 codes as done previously). These specialisations are calculated based on PATSTAT 2021b. The size of the nodes represents the RTA Index value: The larger the node, the higher the relative use of the technology linked to it in the creation of digital sustainability inventions in comparison to its average use in all inventions for each interval. For more information, see Section 1.3 in Part II of this report.

Box 1. USING THE RTA INDEX TO MEASURE SPECIALISATIONS

Balassa (1965) proposed the Revealed Comparative Advantage (RCA) index to measure the comparative advantage of countries in the exporting of manufactured goods. It has been widely applied in the economics literature to understand how selected entities (e.g., countries, companies, industries, etc.) perform in comparison to a relative global average in the output of products. Soete (1987) adapted the specialisation measure to technologies and proposed the Revealed Technology Advantage (RTA) index. We use the RTA index to measure how selected entities perform in relation to a global average. We use the 4-digit CPC codes as the reference for identifying technologies used to create patents. The calculation of the RTA index for a given entity is as follows:

$$\mathrm{RTA}_{\mathrm{CPC}\ code\ c,\mathrm{entity\ e}} = \frac{\dfrac{\text{Number of patents classified with code c linked to entity e}}{\text{Total number of codes linked to entity } e}}{\dfrac{\text{Total number of patents classified with code } c \text{ for all entities}}{\text{Total number of codes linked to for all entities}}}$$

If the RTA index is equal to or above 1, the entity has a specialisation advantage in the respective technology. The "entities" in this report are technologies, firms, or countries. In particular, we use the RTA index for identifying (i) the main technologies used to create digital sustainability inventions (see Figure 3 & Section 1.3 in Part II), (ii) for analysing the existing and missing technological specialisation of R&D Scoreboard companies relevant for the digital sustainability inventions in comparison to a global average (see Figure 9), and (iii) for investigating how R&D Scoreboard companies from specific countries perform in comparison to an average performance of all countries in selected technologies (see Figure 11 and Figure 12).

From the available 670 4-digits CPC codes in the technology space, 62 show at least one specialisation in the considered intervals. We identify four clusters of relevant technologies based on the position of these specialisations in the technology space. These clusters hold 53 specialisations. Knowledge within clusters is more similar than outside it. Since the position of nodes in the technology space is determined by the number of combinations between them, one can say that the technologies identified in the same cluster are more often combined with each other to create inventions than technologies outside the cluster. This does not avoid that technologies from distinct clusters are combined to each other, but highlights that these recombinations are less likely to occur. Combinations within clusters are arguably linked to the emergence of 'related' inventions, whereas combinations with technologies from distinct clusters would potentially lead to more 'unrelated' inventions. Whereas the former is often described as being more common and cheaper, the latter is rare and often linked to the idea of technological breakthroughs (see for example Castaldi et al., 2015).

Therefore, it is relevant to review the particularities of these four clusters and the technologies that compose them. Cluster 1 hold digital sustainability inventions related to energy generation and data-related technologies. We can link 24 CPC codes to this cluster (see Table 3). This cluster takes the most central position in the technology space, besides showing the highest number of specialisations of all four identified clusters across all three intervals (21 in the 1st, 18 in the 2nd, and 17 in the 3rd interval). Cluster 1 includes three sustainability Y02 codes, namely 'Climate change mitigation technologies related to buildings' (Y02B), 'Climate change mitigation technologies in ICT' (Y02D), and 'Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution' (Y02E). The Y02E and Y02B codes, which are linked to energy generation, transmission, or distribution, and end-user applications, respectively, appear strongly linked to CPC codes related to energy generation (e.g., F03D, G21D, and H02S), transmission/distribution (e.g., H04B, H02J, H02M, and Y04S), and inhouse electric technologies (e.g., G01D, F21W, H02H, and H05B).

Table 3: Technologies relevant to the deployment of digital sustainability inventions in Cluster 1.

Cluster No.	CPC code	Description	No. of occur.	Growth 1st - 3rd Interval
1	F03D	Wind motors	1,158	617%
1	F21W	Indexing scheme associated with subclasses F21K, F21L, F21S and F21V, relating to uses or applications of lighting devices or systems	513	2394%
1	G01D	Measuring not specially adapted for a specific variable; arrangements for measuring two or more variables not covered in a single other subclass; tariff metering apparatus; measuring or testing not otherwise provided for	2,359	98%
1	G01W	Meteorology	423	1104%
1	G05D	Systems for controlling or regulating non-electric variables	1,743	922%
1	G05F	Systems for regulating electric or magnetic variables	582	1335%
1	G06F	Electric digital data processing	15,745	235%
1	G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for	13,278	484%
1	G07C	Time or attendance registers; registering or indicating the working of machines; generating random numbers; voting or lottery apparatus; arrangements, systems or apparatus for checking not provided for elsewhere	1,091	695%
1	G08C	Transmission systems for measured values, control or similar signals	1,468	1120%
1	G21D	Nuclear power plant	265	224%
1	H02H	Emergency protective circuit arrangements	736	366%
1	H02J	Circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy	14,188	906%
1	H02M	Apparatus for conversion between ac and ac, between ac and dc, or between dc and dc, and for use with mains or similar power supply systems; conversion of dc or ac input power into surge output power; control or regulation thereof	1,287	316%
1	H02S	Generation of electric power by conversion of infra-red radiation, visible light or ultraviolet light, e.g. Using photovoltaic [PV] modules	1,264	2788%
1	H04B	Transmission	3,367	99%
1	H04L	Transmission of digital information, e.g. Telegraphic communication	13,851	206%
1	H04Q	Selecting	917	75%
1	H04W	Wireless communication networks	13,789	174%
1	H05B	Electric heating; electric light sources not otherwise provided for; circuit arrangements for electric light sources, in general	3,744	903%
1	Y02B	Climate change mitigation technologies related to buildings, e.g. Housing, house appliances or related end-user applications	30,973	786%
1	Y02D	Climate change mitigation technologies in information and communication technologies [ICT], i.e. Information and communication technologies aiming at the reduction of their own energy use	32,829	118%
1	Y02E	Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution	31,885	880%
1	Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. Smart grids	26,601	700%

Source: PATSTAT 2019a.

The Y02D code, which focuses on ICT technologies, appears most strongly linked to data processing and data transmission technologies (e.g., G06F, G06Q, H04L, and H04W). Exemplary inventions of cluster 1 are 'Ecological Bioclimatic System for Supplying Water and Energy in a Housing', 'Method and Sensor Node for Providing Adaptive Sampling in Wireless Sensor Networks, 'Bp Neural Network Photovoltaic Power Prediction Method Based on Information Fusion Theory', and 'GHG Gas Emission Trading System' (see for a detailed description Table C1 in Appendix C).

We can link the second cluster to technologies used for capture, storage, sequestration or disposal of GHGs (code Y02C). Besides being smaller, this specific cluster also disappears over time. It holds only three specialisations in the 1st interval, two in the 2nd, and zero in the 3rd. Linked to it are two codes related to very specific applications of ICT technologies, namely 'Bioinformatics' (G16B) and 'ICT for specifically adapted application fields' (G16Z) (see Table 4). Exemplary patents linked to Cluster 2 are 'Fuel Store Featuring Removal of CO2' and 'Method and System for Managing GHG Emission Quantity in Logistics Processes' (for a detailed description see Table C2 in Appendix C).

Table 4: Technologies relevant to the deployment of digital sustainability inventions in Cluster 2.

Cluster No.	CPC code	Description	No. of occur.	Growth 1st - 3rd Interval
2	G16B	Bioinformatics, i.e. Information and communication technology [ICT] specially adapted for genetic or protein-related data processing in computational molecular biology	138	-11%
2	G16Z	Information and communication technology [ICT] specially adapted for specific application fields, not otherwise provided for	245	418%
2	Y02C	Capture, storage, sequestration or disposal of greenhouse gases [GHG]	90	106%

Source: PATSTAT 2019a.

We can link Cluster 3, in turn, to the processing of goods and domestic applications. It includes three Y02 subclasses, namely 'Technologies for adaptation to climate change' (Y02A), 'Climate change mitigation technologies in the production or processing of goods' (Y02P), 'Climate change mitigation technologies related to wastewater treatment or waste management' (Y02W). Cluster 3 is the second largest cluster and the only that continuously grows over time in terms of the number of specialisations (8 in the 1st interval, 5 in the 2nd, and 11 in the 3rd). It contains the technology most frequently used in the identified patents (Y02P) and shows the largest average growth of individual technologies deployed (see Table 5). Apart from Y02P, we find a number of other technologies directly linked to the production or processing of goods such as 'Working metallic powder' (B22F), 'Additive manufacturing' (B33Y), 'Healthcare informatics' (G16H), and 'Superheating of steam' (F22G). In turn, we find a number of technologies that can broadly summarize under specific domestic application, which include apart from Y02A and Y02W technologies such as 'Solar heat collectors or systems' (F24S), 'Gathering or removal of domestic or like refuse' (B65F), 'Installations or methods for obtaining, collecting, or distributing water' (E03B), and 'Domestic- or space-heating systems' (F24D). Examples of patents linked to Cluster 3 are 'Real Time Energy Consumption Analysis and Reporting', 'Waste Management System', 'Technique for Determining and Reporting Reduction in Emissions of GHG at a Site', and 'Curtain Wall Window Structure and Full Daylight Solar Air-Conditioner' (for a detailed description see Table C3 in Appendix C).

Table 5: Technologies relevant to the deployment of digital sustainability inventions in Cluster 3.

Cluster No.	CPC code	Description	No. of occur.	Growth 1st - 3rd Interval
3	B22F	Working metallic powder; manufacture of articles from metallic powder; making metallic powder; apparatus or devices specially adapted for metallic powder	1,171	5989%
3	B25J	Manipulators; chambers provided with manipulation devices	678	283%
3	B33Y	Additive manufacturing, i.e. Manufacturing of three-dimensional [3-D] objects by additive deposition, additive agglomeration or additive layering, e.g. By 3-D printing, stereolithography or selective laser sintering	1,540	4869%
3	B65F	Gathering or removal of domestic or like refuse	404	888%
3	E03B	Installations or methods for obtaining, collecting, or distributing water	356	2655%
3	F22G	Superheating of steam	78	Inf (from 0 to 78)
3	F24D	Domestic- or space-heating systems, e.g. Central heating systems; domestic hot-water supply systems; elements or components therefor	438	597%
3	F24S	Solar heat collectors; solar heat systems	604	626%
3	G05B	Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements	15,796	318%
3	G16H	Healthcare informatics, i.e. Information and communication technology [ICT] specially adapted for the handling or processing of medical or healthcare data	909	50%
3	Y02A	Technologies for adaptation to climate change	12,222	660%
3	Y02P	Climate change mitigation technologies in the production or processing of goods	49,391	217%

3	Y02W	Climate change mitigation technologies related to wastewater treatment or waste management	2,991	407%

Source: PATSTAT 2019a.

Cluster 4, in turn, is particularly linked to transportation technologies. It has only one Y02 subclass, namely 'Climate change mitigation technologies related to transportation' (Y02T) and is the third largest. This cluster slightly shrinks in terms of the number of technologies with a specialisation over time (9 in the 1st, and 7 in the 2nd and 3rd interval). An overview of the codes linked to this cluster is presented in Table 6. There is a variety of technologies linked to vehicles and engines for civilian and military use (see Table 6), such as 'Propulsion of electrically-propelled vehicles' (B60L), 'Power supply lines' (B60M), 'Conjoint control of vehicle sub-units' (B60W), 'Controlling combustion engines' (F02D), 'Locomotives' (B61C), 'Offensive or defensive arrangements on vessels' (B63G), 'Aeroplanes; helicopters' (B64C), 'Ground or aircraft-carrier-deck installations' (B64F), or 'Indexing scheme relating to wind, spring, weight, inertia or like motors, to machines or engines for liquids' (F05B). Examples of patents linked to Cluster 4 are 'Energetically-Autonomous Transportation Vehicle Using Multiple Green Energy Sources' and 'New Energy Rail Bus Transit System' (for a detailed description see Table C4 in Appendix C).

Table 6: Technologies relevant to the deployment of digital sustainability inventions in Cluster 4.

Cluster No.	CPC code	Description	No. of occur.	Growth 1st - 3rd Interval
4	B60L	Propulsion of electrically-propelled vehicles; supplying electric power for auxiliary equipment of electrically-propelled vehicles; electrodynamic brake systems for vehicles in general; magnetic suspension or levitation for vehicles; monitoring operating variables of electrically-propelled vehicles; electric safety devices for electrically-propelled vehicles	7,663	849%
4	B60M	Power supply lines, and devices along rails, for electrically-propelled vehicles	118	1300%
4	B60W	Conjoint control of vehicle sub-units of different type or different function; control systems specially adapted for hybrid vehicles; road vehicle drive control systems for purposes not related to the control of a particular sub-unit	1,351	263%
4	B60Y	Indexing scheme relating to aspects cross-cutting vehicle technology	597	1422%
4	B61C	Locomotives; motor railcars	71	107%
4	B62H	Cycle stands; supports or holders for parking or storing cycles; appliances preventing or indicating unauthorized use or theft of cycles; locks integral with cycles; devices for learning to ride cycles	94	1750%
4	B63G	Offensive or defensive arrangements on vessels; mine-laying; mine-sweeping; submarines; aircraft carriers	39	200%
4	B64C	Aeroplanes; helicopters	561	516%
4	B64F	Ground or aircraft-carrier-deck installations specially adapted for use in connection with aircraft; designing, manufacturing, assembling, cleaning, maintaining or repairing aircraft, not otherwise provided for; handling, transporting, testing or inspecting aircraft components, not otherwise provided for	155	808%
4	F02D	Controlling combustion engines	930	4%
4	F03B	Machines or engines for liquids	275	767%
4	F05B	Indexing scheme relating to wind, spring, weight, inertia or like motors, to machines or engines for liquids covered by subclasses F03B, F03D and F03G	825	367%
4	Y02T	Climate change mitigation technologies related to transportation	16,468	583%

Source: PATSTAT 2019a.

To summarize, Section 2 of the report helps our main understanding of the nature and evolution of digital sustainability inventions. Considering patent-based indicators for the period from 2001 to 2018, we suggest that the core of digital sustainability technologies is linked to 'Climate change mitigation technologies related to transportation' (Y02T), 'Climate change mitigation technologies in ICT' (Y02D), 'Climate change mitigation technologies related to buildings' (Y02B), Reduction of GHG emissions, related to energy generation, transmission or distribution' (Y02E), as well as 'Climate change mitigation technologies in the production or processing of goods' (Y02P), which are most frequently used to create digital sustainability inventions. In terms of the occurrence, we find the majority of digital sustainability priorities patents linked to 'Climate change mitigation technologies in the production or processing of goods' (Y02P).

Yet such digital sustainability priorities inventions are relatively often restricted to the jurisdiction of the initial (domestic) patent office, whereas 'Climate change mitigation technologies in ICT' (Y02D) are diffusing more internationally.

Furthermore, we identify four main clusters of related technologies, which are linked to the creation of particular digital sustainability inventions during the period from 2001 to 2018: (1) Energy generation and data-related technologies, (2) technologies related to the capture, storage, sequestration or disposal of greenhouse gases, (3) technologies related to the processing of goods and domestic applications, and (4) technologies related to transportation. The four clusters together hold 53 specialised technologies. Over time, we find that Cluster 1 unites the highest number of specialized technologies. Both cluster 3 and 4 have been increasing the number of specialised technologies, whereas Cluster 2 has been losing specialised technologies over time. Thus, we can detect very distinct patterns in these clusters of related technologies, which create distinct digital sustainability inventions. At this stage, the growth trajectory of digital sustainability inventions seems strongest for Cluster 3, which has technologies related to the processing of goods and domestic applications.

3 DIGITAL SUSTAINABILITY TECHNOLOGIES BY R&D SCOREBOARD COMPANIES

In the next step of the analysis, we investigate the involvement of the world's top corporate R&D investors in the development of digital sustainability inventions. We apply the proposed patent identification strategy to PATSTAT 2021b. This update not only increases the number of digital sustainability patents identified from 319,243 (using PATSTAT 2019a) to 494,255, but also allows improved coverage of more recent years. We identify priority patents using the IP5 strategy for the period from 2016 to 2018 (see Part II of this report, Section 3). Using the IP5 strategy excludes patents filled in just one patent office and generates a dataset of related inventions with a 'comparable' technological and economic value (Dernis et al., 2015). Given that R&D Scoreboard companies have large, complex and changing ownership structures (due to firm exits, M&As, etc.), we need to restrict the firm level analysis of digital sustainability patenting to a relatively short period of 3 years for which we assume ownership structures (ibid). We link patents to R&D Scoreboard companies using the JRC-OECD COR&DIP© v.3 dataset (Hernández et al. (2020); Amoroso et al. 2021).

In the subsequent analysis, we first investigate the output of digital sustainability inventions by R&D Scoreboard companies (Subsection 3.1). Thereby, we consider their global contribution in comparison to other actors as well as a breakdown of output of digital sustainability patents by R&D Scoreboard companies by country, sector, industry, and individual firm level. Second, we analyse the technological specialisation of R&D Scoreboard companies in the technologies identified as relevant to the deployment of digital sustainability inventions (Subsection 3.2). Thereby, we consider the technological specialisations of all R&D Scoreboard companies jointly and the specialisation of individual firms. Finally, we focus on the specialisation pattern observed for the European R&D Scoreboard companies.

3.1 Output

3.1.1 Global contribution

We would like to develop a better understanding of the contribution of R&D Scoreboard companies in the global production of all digital sustainability priorities. Therefore, we compare their output to other actors, which include all other corporate actors, universities, government, or individuals. Since we use the IP5 strategy, we consider only priorities registered at one of the five largest IP offices AND registered in at least two distinct patent offices worldwide. As universities (Veer and Jell, 2012) and governments arguably do not extend their IP rights to other patent offices, their presence in our sample of other actors is probably low. Applying the IP5 identification strategy to PATSTAT 2021b, we find a total of 537,069 priorities for the period 2016 to 2018. R&D Scoreboard companies account for 325,508 (60.6%) of all priorities (see Table 7).

Table 7: Number of priorities and digital sustainability priorities for R&D Scoreboard and non-Scoreboard actors (20016-2018).

	No. of priorities by R&D Scoreboard companies	No. of priorities by other actors	Share of R&D Scoreboard companies
All priorities	325,508	211,561	60.6%
Digital sustainability priorities	5,057	2,751	64.8%

Source: PATSTAT 2021b.

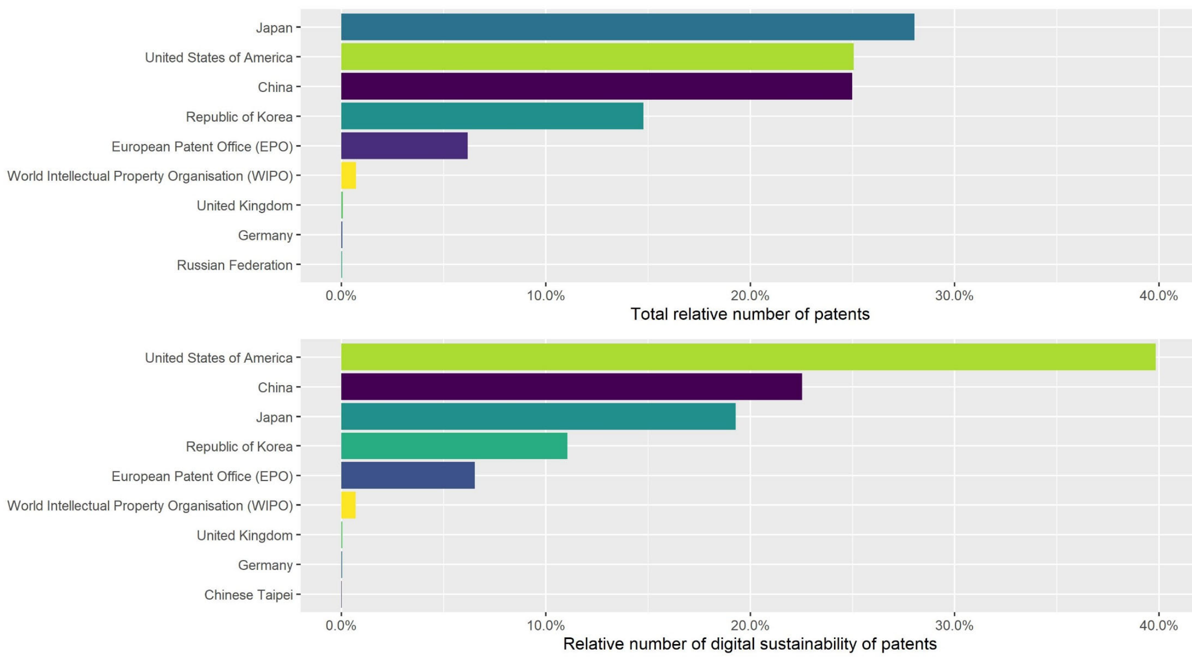
53,744 priorities (10.01% of all priorities) identified in the considered period are classified as 'Technologies or applications for mitigation or adaptation against climate change' (code Y02). 35,835 of these (66.7%) are owned by R&D Scoreboard companies. Considering digital sustainability patents, there is a total of 7,808 priorities registered in the considered period, from which 5,057 (64.8%) are owned by scoreboard companies. 'Other actors' account globally for 40.4% of all patents, and 35,2% of all digital sustainability priorities (see Table 7). Thus, for the observation period, only 1.45% (7,808/537,069) of all priority filings and 14.53% (7,808/53,744) of all sustainable priorities3 can be considered as digital sustainability inventions. For R&D Scoreboard companies, these figures are roughly similar: 1.55%

[3] For comparison, Amoroso et al. (2021) report almost 20% (average across different sectors) of the Y02 priorities as digital sustainability patents. Yet, the report employs exclusively the ICT classification (Inaba and Squicciarini, 2017), rather than a mix of search modules, to identify 'digital patents' in the identification strategy for digital sustainability patents.

(5,057/325,508) and 14.11% (5,057/35,835) of all R&D Scoreboard companies' priority filings and sustainable priority filings are considered digital sustainability priorities, respectively.

Considering priorities by R&D Scoreboard companies and other actors jointly for the 3-year observation period, we find priorities from nine different patent offices (see Figure 4). Since we employ the IP5 identification strategy, the patent offices from Japan, the US, China, South Korea, and the European Patent Office (EPO) dominate the sample. It should be noted that the very high number of priorities for Japan and China have previously been linked to a phenomenon known as 'patent flooding', where actors file numerous patents with minor changes around the core technology of a patent they own (Wolfson, 1993). Therefore, it is more appropriate to consider differences between priorities in general and digital sustainability priorities for each patent office. From this perspective, we observe proportionally higher shares of digital sustainability patents compared to all patents for patent offices from the US and China.

Figure 4: Shares of priorities and digital sustainability priorities of all actors by patent offices (2016-20818).

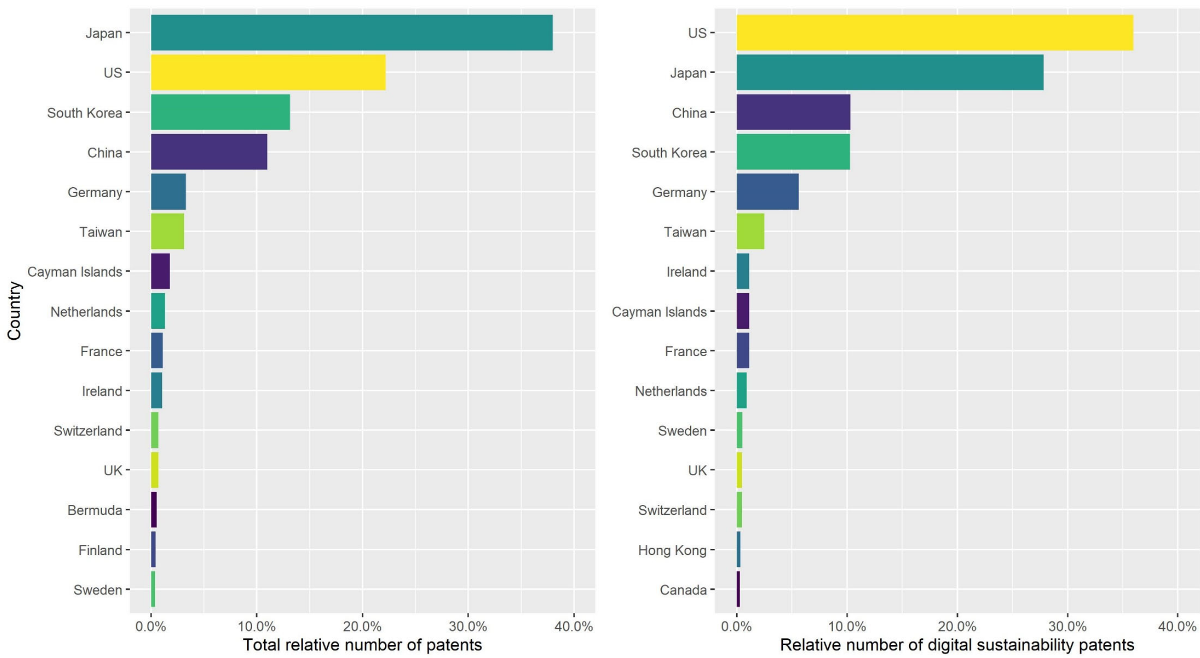


Source: PATSTAT 2021b.

3.1.2 Geography, sector and industry

In the following analysis of the output of digital sustainability priorities, we exclude other actors and consider only R&D Scoreboard companies. Within this sample, we find 40 different countries locating headquarters of R&D Scoreboard companies (with at least one priority patent). However, R&D Scoreboard companies from only 25 countries have at least one digital sustainability priority. Considering the top 15 countries regarding total number of priorities and total number of digital sustainability priorities (Figure 5), we observe R&D Scoreboard headquarters being mostly located in the US, Japan, China, and South Korea in both categories. R&D Scoreboard companies from the following countries patent proportionally more in digital sustainability patents compared to their relative share of all patents during the observation period: US, Germany, Ireland, France, Sweden, Hong Kong and Canada.

Figure 5: Top 15 countries of R&D Scoreboard companies for share in total priorities and digital sustainability priorities (2016-2018).



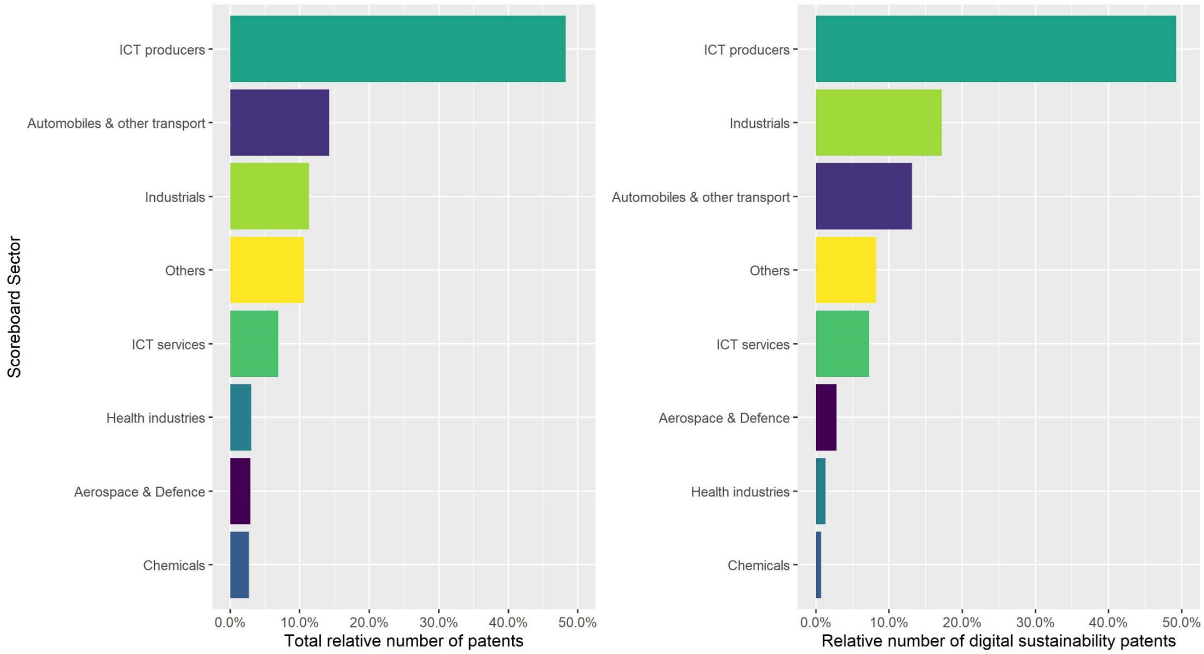
Source: PATSTAT 2021b.

Next, we analyse the share of total and digital sustainability priorities for R&D Scoreboard companies in eight main economic sectors. The economic sectors considered are the eight proposed in Hernández et al. (2018) when analysing R&D Scoreboard companies. They are partially based on the Industry Classification Benchmark (ICB) (see Appendix D). As these sectors were defined specifically for analysing R&D Scoreboard companies, we refer to them as 'Scoreboard sectors'.

We observe that R&D Scoreboard companies in the Scoreboard sector of 'ICT producers' lead in terms of the relative the share total as well as digital sustainability patenting during the observation period (see Figure 6). The shares are 48.23% (156,993 priorities) and 49.27% (2,492 priorities), respectively. The other top Scoreboard sectors in digital sustainability patenting are 'Industrials' (17.22%), 'Automobiles & other transport' (13.16%), 'Others' (8.22%), and 'ICT Services' (7.29%). We find that R&D Scoreboard companies in 'Industrials' patent proportionally more in digital sustainable technologies than their share in total patenting would suggest. 'ICT producers' and 'ICT services' also do it, by a smaller extent (1.04% and 0.36% more, respectively, compared to 5.88% for 'Industrials'). R&D Scoreboard companies in 'Others', 'Chemicals', 'Health industries', 'Automobiles & other transport', and 'Aerospace & Defence' patent proportionally less in digital sustainability technologies compared to their respective shares in general patenting (namely 2.38%, 2.01%, 1.73%, 1.08%, and 0.09% less, respectively).

Figure 6:

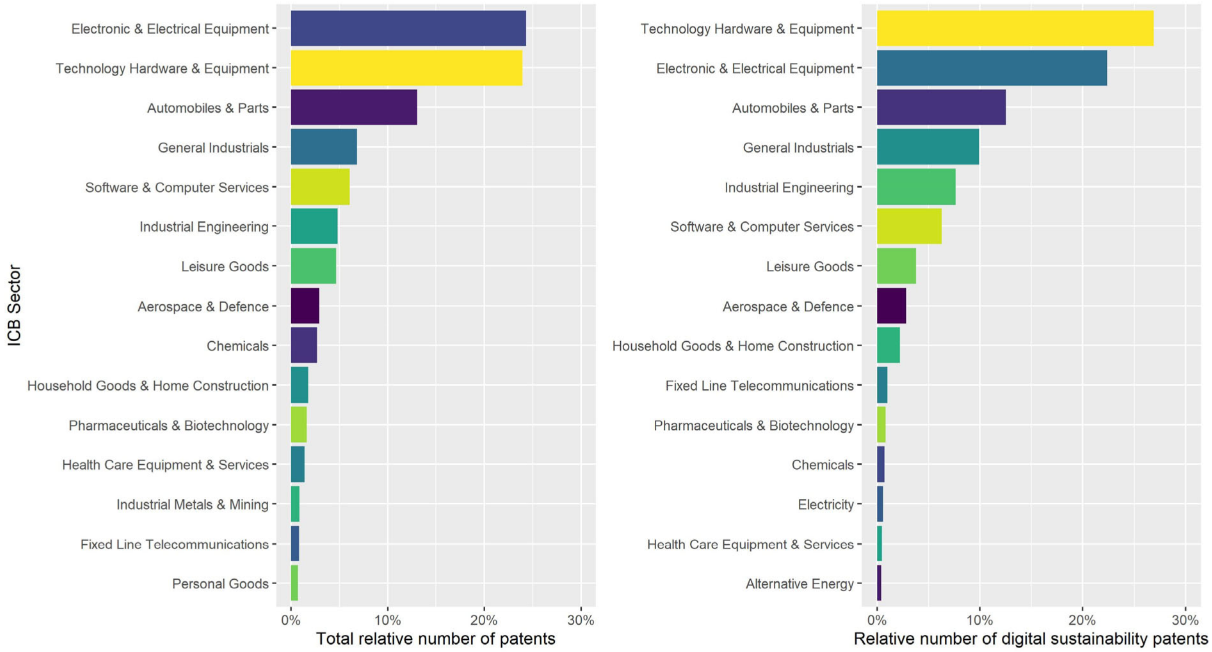
Figure 6: Share of total and in digital sustainability priorities for R&D Scoreboard companies by Scoreboard sectors.



Source: PATSTAT 2021b.

We can further disaggregate this Scoreboard sectors into their original ICB sectors. The R&D Scoreboard companies are linked to 37 ICB sectors, of which 27 have at least one digital sustainability patent registered during the observation period. Considering the top ICB sectors in digital sustainability priorities, we find 'Technology Hardware & Equipment' in the lead (26.90%), followed by 'Electronic & Electrical Equipment' (22.37%), 'Automobiles & Parts' (12.53%), 'General Industrials' (9.92%), and 'Industrial Engineering' (7.62%) (see Figure 7). We find that R&D Scoreboard companies in 'General Industrials', 'Technology Hardware & Equipment', and 'Industrial Engineering' patent proportionally more in digital sustainable technologies than their share in total patenting would suggest (3.11%, 2.98%, 2.80% higher shares in digital sustainability patents than in general patents, respectively). This applies also, to a smaller extent, to R&D Scoreboard companies in 'Electricity', 'Household Goods & Home Construction', 'Alternative Energy', 'Software & Computer Services', and 'Fixed Line Telecommunications' (0.43%, 0.39%, 0.35%, 0.23%, and 0.15% more, respectively). Conversely, the ICB sectors 'Chemicals', 'Electronic & Electrical Equipment', 'Health Care Equipment & Services', 'Leisure Goods', 'Pharmaceuticals & Biotechnology', 'Personal Goods, Automobiles & Parts', 'Industrial Metals & Mining', and 'Aerospace & Defence' patent less in digital sustainability technologies than their general patenting activity would suggest (namely 2.01%, 1.94%, 0.91%, 0.87%, 0.82%, 0.71%, 0.61%, 0.60%, and 0.09% less, respectively).

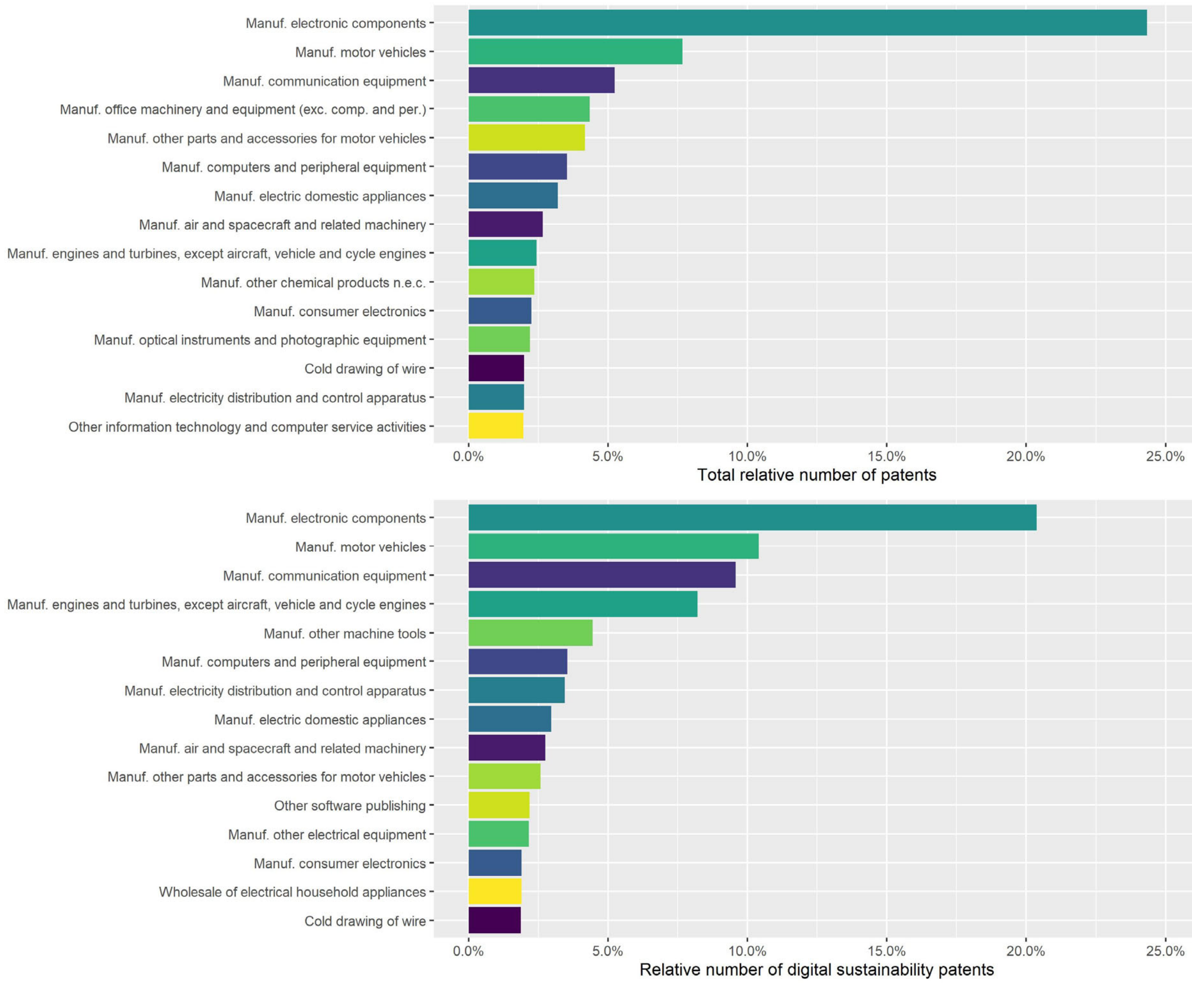
Figure 7: Share of total and in digital sustainability priorities for R&D Scoreboard companies by ICB sectors.



Source: PATSTAT 2021b.

Finally, we analyse the distributions of patenting of R&D Scoreboard companies by NACE industries. Out of the 289 NACE industries (4-digit level) linked to R&D Scoreboard companies, we find companies of 93 industries (32%) that have at least one digital sustainability patent. For better visualization, we focus on the top 15 NACE industries in terms of the share of total and in digital sustainability priorities (see Figure 8).

Figure 8: Share of total and in digital sustainability priorities for R&D Scoreboard companies by Top 15 NACE industry (4-digit).



Source: PATSTAT 2021b.

'Manufacturing of electronic components', 'Manufacturing of motor vehicles', and 'Manufacturing of communication equipment' lead in general patenting as well as patenting in digital sustainability. R&D Scoreboard companies from 'Manufacturing of electronic components' patented 24.34% and 20.37%, respectively, of the patents from these categories. R&D Scoreboard companies in 'Manuf. engines and turbines, except aircraft, vehicle and cycle engines', 'Manuf. communication equipment', 'Manuf. other machine tools', 'Manuf. motor vehicles', 'Manuf. electricity distribution and control apparatus', 'Manuf. other electrical equipment', 'Wholesale of electrical household appliances', 'Other software publishing', 'Manuf. air and spacecraft and related machinery', and 'Manuf. computers and peripheral equipment' patent more in digital sustainable technologies than their share in total patenting would suggest (5.77%, 4.34%, 3.81%, 2.73%, 1.45%, 0.65%, 0.48%, 0.29%, 0.09%, and 0.01% more, respectively). Conversely, companies in 'Manuf. office machinery and equipment (exc. comp. and per.)', 'Manuf. electronic components', 'Manuf. other chemical products n.e.c.', 'Manuf. optical instruments and photographic equipment', 'Other information technology and computer service activities', 'Manuf. other parts and accessories for motor vehicles', 'Manuf. consumer electronics', 'Manuf. electric domestic appliances', and 'Cold drawing of wire' patent proportionally less (4.35%, 3.97%, 2.36%, 2.19%, 1.96%, 1.59%, 0.35%, 0.23%, 0.12% less, respectively).

3.1.3 Firm level

Next, we disaggregate individual firm data to identify the most important R&D Scoreboard companies in terms of digital sustainability patenting. We find that only 403 (20.2%) of the 2,000 scoreboard companies have at least one digital sustainability priority filed between 2016 and 2018. We present the Top 50 R&D Scoreboard companies regarding number of digital sustainability patents here (see Table 8). 27 out of these 50 R&D Scoreboard companies are also among the Top 50 R&D Scoreboard companies in CCMA technologies presented in Amoroso et al. 2021 (see Amoroso et al. 2021, Table 4.1., p. 56). These overlapping companies are highlighted in green in Table 8.

Table 8: Top 50 R&D Scoreboard companies in terms of digital sustainability priority patents (2016 - 2018).

Name (abbreviated)	**Country**	**ICB Sector**	**Nace Rev2 code primary**	**Number of Digital Sustain. Priorities**	**Share**	**Rank**
Qualcomm	US	Technology Hardware & Equipment	2630	236	4.6%	1
Intel	US	Technology Hardware & Equipment	2611	230	4.5%	2
General Electric	US	General Industrials	2811	223	4.4%	3
Samsung Electronics	South Korea	Electronic & Electrical Equipment	2611	213	4.2%	4
Fanuc	Japan	Industrial Engineering	2849	168	3.3%	5
Siemens	Germany	Electronic & Electrical Equipment	2811	164	3.2%	6
Ford Motor	US	Automobiles & Parts	2910	119	2.3%	7
Hitachi	Japan	Electronic & Electrical Equipment	2712	104	2.0%	8
Zte	China	Technology Hardware & Equipment	2630	96	1.9%	9
Huawei Investment & Holding Co	China	Technology Hardware & Equipment	4643	95	1.9%	10
Omron	Japan	Electronic & Electrical Equipment	2611	89	1.8%	11
Microsoft	US	Software & Computer Services	5829	77	1.5%	12
Huagong Tech	China	Electronic & Electrical Equipment	2790	77	1.5%	13
Panasonic	Japan	Leisure Goods	2751	76	1.5%	14
Toyota Motor	Japan	Automobiles & Parts	2910	76	1.5%	15
LG Electronics	South Korea	Leisure Goods	2640	70	1.4%	16
Toyota Industries	Japan	Automobiles & Parts	2910	67	1.3%	17
Honeywell	US	General Industrials	2932	58	1.1%	18
Boeing	US	Aerospace & Defence	3030	57	1.1%	19
United Technologies	US	Aerospace & Defence	3030	50	1.0%	20
Mediatek	Taiwan	Technology Hardware & Equipment	2611	49	1.0%	21
Canon	Japan	Technology Hardware & Equipment	2823	47	0.9%	22
Sumitomo Electric	Japan	Electronic & Electrical Equipment	2434	47	0.9%	23
Denso	Japan	Automobiles & Parts	2932	46	0.9%	24
Yokogawa Electric	Japan	Electronic & Electrical Equipment	2829	45	0.9%	25
Kia Motors	South Korea	Automobiles & Parts	2910	43	0.8%	26
Advanced Micro Devices	US	Technology Hardware & Equipment	2611	42	0.8%	27
Haier Smart Home	China	Household Goods & Home Construction	2751	41	0.8%	28
Emerson Electric	US	Electronic & Electrical Equipment	2651	39	0.8%	29
Alphabet	US	Software & Computer Services	6209	39	0.8%	30
General Motors	US	Automobiles & Parts	2910	38	0.7%	31
BMW	Germany	Automobiles & Parts	2910	37	0.7%	32
LS	South Korea	General Industrials	2434	34	0.7%	33
Fujitsu	Japan	Software & Computer Services	2620	33	0.6%	34
Apple	US	Technology Hardware & Equipment	2620	33	0.6%	35
IBM	US	Software & Computer Services	6201	32	0.6%	36
Seiko Epson	Japan	Technology Hardware & Equipment	2611	32	0.6%	37
Mitsubishi Heavy	Japan	General Industrials	2899	32	0.6%	38
Honda Motor	Japan	Automobiles & Parts	3091	31	0.6%	39
Hyundai Motor	South Korea	Automobiles & Parts	2910	30	0.6%	40
Midea Group	China	Household Goods & Home Construction	4754	29	0.6%	41
Boe Technology Group	China	Electronic & Electrical Equipment	2611	29	0.6%	42
LG Chem	South Korea	General Industrials	2059	29	0.6%	43

Metallurgical of China	China	General Industrials	4120	27	0.5%	44
Schneider	France	Electronic & Electrical Equipment	2712	27	0.5%	45
Micron Technology	US	Technology Hardware & Equipment	2611	27	0.5%	46
Nio	Cayman Islands	Automobiles & Parts	2910	26	0.5%	47
Johnson Controls	Ireland	General Industrials	2825	26	0.5%	48
Nari Technology	China	Software & Computer Services	6201	26	0.5%	49
Fujifilm	Japan	Electronic & Electrical Equipment	2670	25	0.5%	50

Source: PATSTAT 2021b.

We can see that the Top 3 R&D companies, namely Qualcomm, Intel, and General Electric, come all from the US. In the fourth position, in terms of patenting in digital sustainability technologies, comes Samsung Electronics from South Korea. Each of these four companies filed more than 200 digital sustainability priorities during the observation period. In the following group of R&D Scoreboard companies with more than 100 digital sustainability priority filings, we have Fanuc (JP), Itachi (JP), Siemens (DE), and Ford Motor (US). From the Top 50 R&D companies, the majority comes from the US and Japan (15 each), followed by China (8), South Korea (6), Germany (2), and Cayman Islands, France, Ireland, and Taiwan tied (1 representative company each). In terms of distribution across ICB sectors, 11 R&D Scoreboard companies are in 'Electronic & Electrical Equipment' sector, 10 in 'Automobiles & Parts' as well as 'Technology Hardware & Equipment', and 7 in 'General Industrials', 5 in 'Software & Computer Services', 2 in 'Aerospace & Defence', 'Household Goods & Home Construction', and 'Leisure Goods', and 1 in 'Industrial Engineering'. For Scoreboard sectors, 21 companies are 'ICT producers', 10 are 'Automobiles & other transport', 8 are 'Industrials', 5 'ICT services', 4 'Others', and 2 'Aerospace & Defence'.

<u>To summarize</u>, our findings for the output of digital sustainability patenting by R&D Scoreboard companies (2018 – 2019), we highlight that only 1.55% of all priority patents filed by R&D Scoreboard companies, and 14.11% of all their sustainable priority patents, are in digital sustainability technologies. Thus, digital sustainability inventions as such as the integration of CCMA and digital technologies is still in an early phase. However, R&D Scoreboard companies make a substantial contribution to this process as they account for 64.8% for digital sustainability priorities, which is proportionally higher than their share in total patenting.

Out of 40 countries, only 25 host R&D Scoreboard companies with digital sustainability inventions. The Top 5 are the US, Japan, China, South Korea, and Germany, whereby the US and China contribute proportionally more compared to their share in total inventions. Amongst the Top 15, we also find France, the Netherlands, Sweden, UK, and Switzerland. Only about 20% of all 2,000 R&D Scoreboard companies have digital sustainability inventions. The Top 50 and Top 10 R&D Scoreboard companies account for 66.2% and 32.3%, respectively, and originate primarily from the US, Japan, China, and South Korea. Amongst the Top 50 R&D Scoreboard, we find only four European companies: Siemens (DE), BMW (DE), Schneider (FR), and Johnson Controls (IE).

R&D Scoreboard companies in 'ICT producers', 'Industrials', and 'Automobile and other transports' produce most digital sustainability inventions. The vast majority, however, is concentrated amongst 'ICT producers'. This refers to the industries of 'Technology Hardware & Equipment' and 'Electronic & Electrical Equipment'. Thus, the core of digital sustainability inventions originates from digital companies embracing sustainable technologies, rather than the other way around. This points to the importance of digital and ICT industries to advance the twin transition.

3.2 Technological specialisations

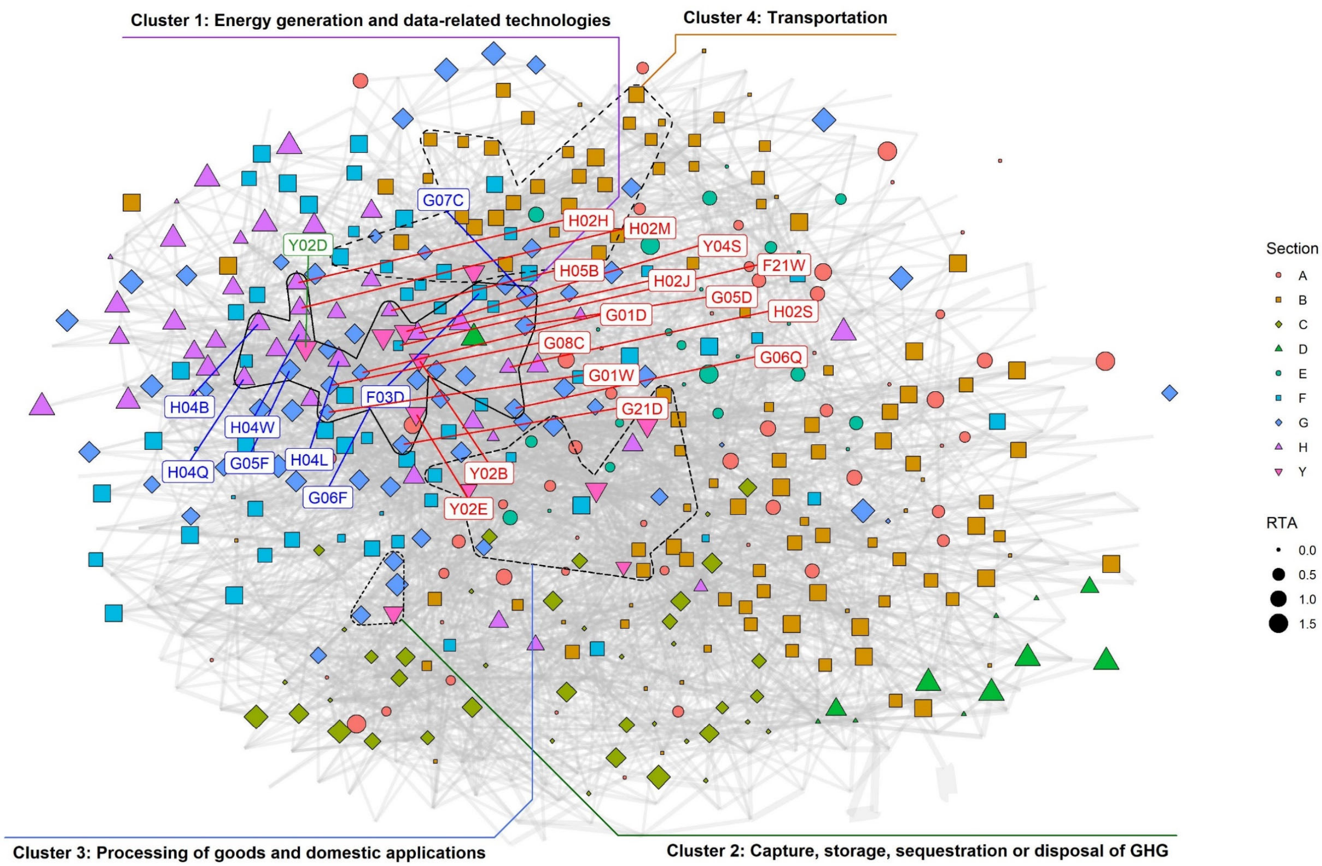
This section considers the specialisation of R&D Scoreboard companies in technologies relevant to the deployment of digital sustainability inventions as previously identified in Section 2 of this report. The premise is that, due to knowledge commonalities (see for example Breschi et al., 2003), having a specialisation in these relevant technologies indicates that R&D Scoreboard companies command the required knowledge to deploy digital sustainability inventions, even if digital sustainability inventions might not yet materialise. We assume that it is easier for R&D Scoreboard companies to explore one specific cluster of related technologies than to explore several distinct clusters. Therefore, the subsequent analysis considers all priority patents registered by R&D Scoreboard companies (i.e., not only digital sustainability priorities) for the observation period (2016-2018). First, we consider available and missing specialisation of all R&D scoreboard companies in the 53 technologies identified as relevant to the deployment of digital sustainability inventions. We differentiate this analysis also according to the geographic origin of R&D Scoreboard companies. Second, we analyse the technological specialisations relevant to the deployment of digital sustainability inventions in the four clusters at the level of the individual R&D Scoreboard companies. Finally, we discuss more in detail the existing and missing technological specialisations of European R&D Scoreboard companies across the four considered clusters.

3.2.1 All R&D Scoreboard companies

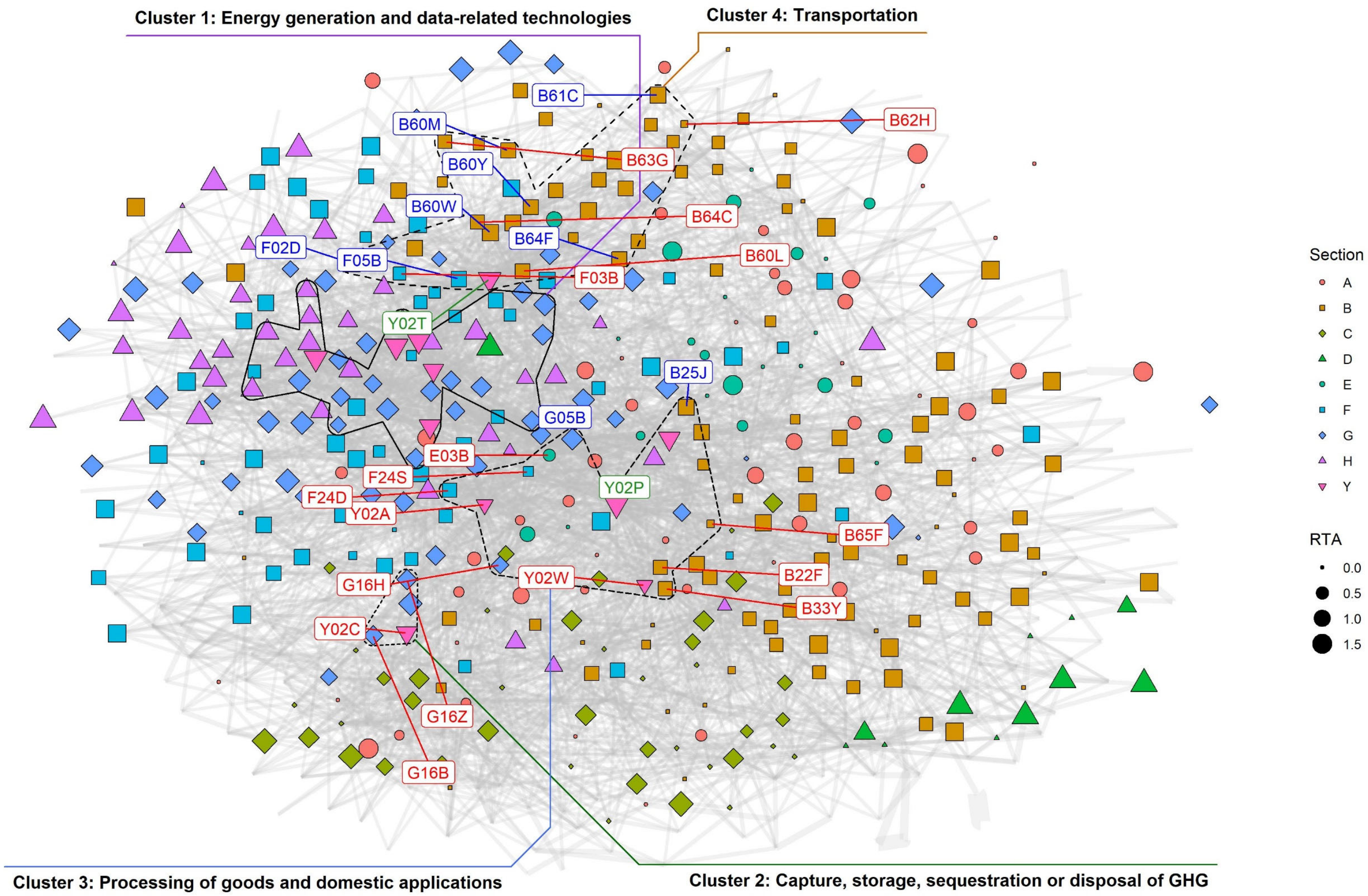
First, we highlight available and missing technological specialisation of all R&D Scoreboard companies in the 53 technologies identified as relevant to the deployment of digital sustainability inventions. We use again a technology space approach but with specialisations calculated for the period 2016 to 2018 (see Figure 9). For better visualization, we present the larger Cluster 1 (Figure 9b) separately from the other three (Figure 9b). We indicate available specialisations in blue and green (in the case of Y02 codes), and missing specialisations in red. R&D Scoreboard companies possess specialisations in three subsections of the Y02 codes, namely 'Climate change mitigation technologies in information and communication technologies' (Y02D) from Cluster 1, 'Climate change mitigation technologies in the production or processing of goods' (Y02P) from Cluster 3, and 'Climate change mitigation technologies related to transportation' (Y02T) from Cluster 4. It seems that R&D Scoreboard companies are well equipped with existing knowledge in the 'transportation cluster' (Cluster 4), with developed specialisations in eight out of the 13 relevant technologies (61.5%).

Figure 9: Existing and missing technological specialisations of R&D Scoreboard companies in Clusters 1 (a), 2, 3, and 4 (b) (2016-2018).

a)



b)



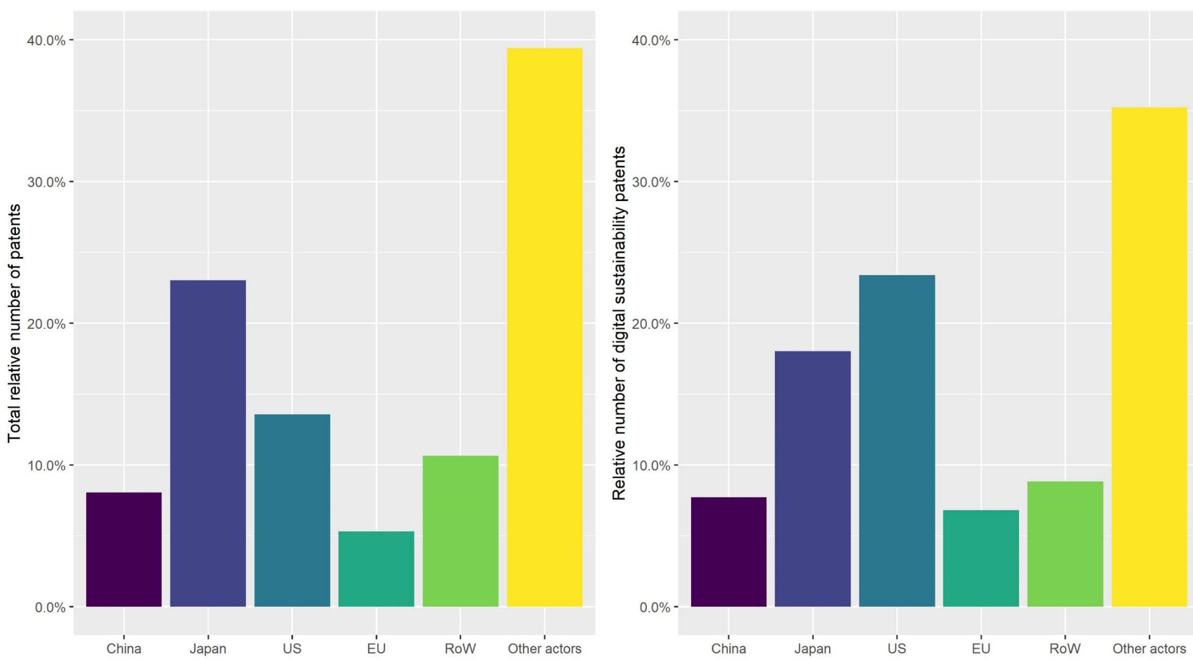
Sources: PATSTAT 2019 (for defining the technology space) and PATSTAT 2021 (for calculating specialisations).

Notes: Based on RTA index for all R&D Scoreboard companies. Blue and green (Y02 subsections) labels indicate specialisation and red labels missing ones. The node size represents the strength of the specialisation: The larger the node, the higher the relative use of the technology by Scoreboard companies in comparison to its average usage by all possible actors.

R&D Scoreboard companies possess also suitable knowledge linked to 'energy generation and data-related technologies' (Cluster 1). In this case, existing specialisations are mostly concentrated on the left side of the cluster, in codes from the electricity section (Section H). It is also on this side of the cluster where 'Climate change mitigation technologies in ICTs' (Y02D) are located. The specialisations of R&D Scoreboard companies into technologies such as 'Electrical digital data processing' (G06F), 'Transmission of digital information' (H04L), and 'Wireless communication' (H04W) indicate that they are particularly well equipped to deploy relevant ICT-related sustainable inventions. The analysis shows that R&D Scoreboard companies as a group do not possess strong specialisations in technologies in the cluster related to 'processing goods and domestic applications' (Cluster 3), and no specialisations in the cluster related to 'capture, storage, sequestration or disposal of GHG' (Cluster 2).

Next, we differentiate the previous analysis according to the location of the headquarters of R&D Scoreboard companies. We focus on the four leading geographies[4] regarding number of digital sustainability patents owned by scoreboard companies, namely China, Japan, the US and the EU27[5]. The patents registered by the remaining geographies linked to R&D Scoreboard companies are presented in the Rest of the World (RoW) category. Although Japanese R&D Scoreboard companies have the highest relative share of all priorities in the observation period, US based R&D Scoreboard companies lead in digital sustainability patents (see Figure 10). In addition, R&D Scoreboard companies based in the EU contribute proportionally more to digital sustainable patenting than general patenting. This does not apply to South Korea and China.

Figure 10: Relative share of all priorities and digital sustainability priorities by geography (2016-2018).



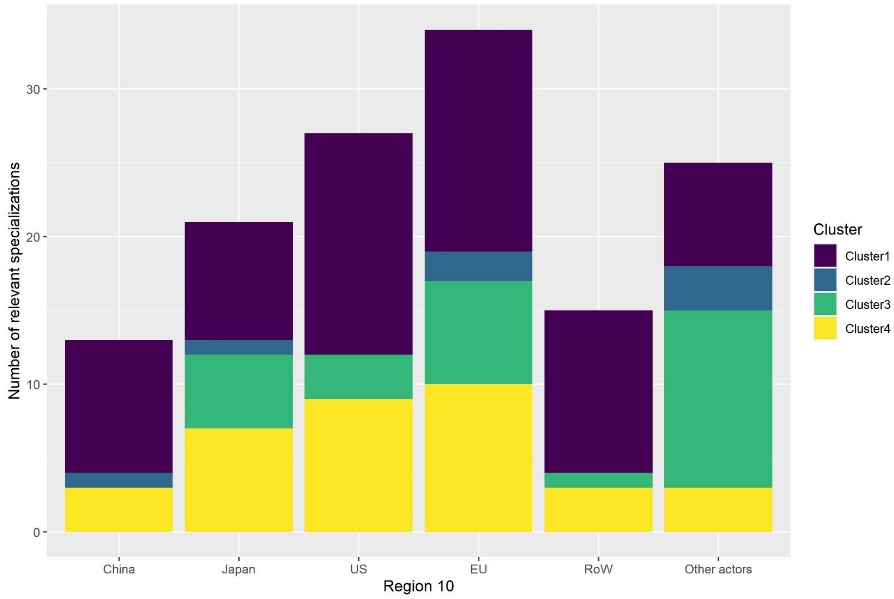
Note: Non-Scoreboard owners are defined as every actor that is not identified as R&D Scoreboard company.
Source: PATSTAT 2021b.

We also consider the number of specialisations by the companies form the respective regions in comparison to the specialisations of other actors in the technologies that compose the four identified clusters (Figure 11). EU R&D Scoreboard companies lead in absolute number of specialisations in relevant technologies, followed by the US.

[4] Additionally, it is noted that South Korea follows as the fifth most relevant geography regarding number of digital sustainability patents owned by Scoreboard companies. Scoreboard companies from the EU27 own 6.81% of all digital sustainability patents, compared to 6.66% owned by South-Korean Scoreboard companies. Thereby, Scoreboard companies from South Korea own the majority of digital sustainability patents allocated to the RoW category presented in Figure 10: 75.42% of all digital sustainability patents linked to the RoW category in this figure (i.e., 6.66% of the 8.83% share linked to Scoreboard companies from the RoW) are owned by Scoreboard companies from South Korea.

[5] EU27 is based on the 2020 update of the classification, i.e., United Kingdom is not included.

Figure 11: Number of specialisations of R&D Scoreboard companies by region (2016-2018).



Notes: Number of specialisations of selected regions in comparison to the number of specialisations held by other actors in the identified relevant technologies for the observation period.
Source: PATSTAT 2021b.

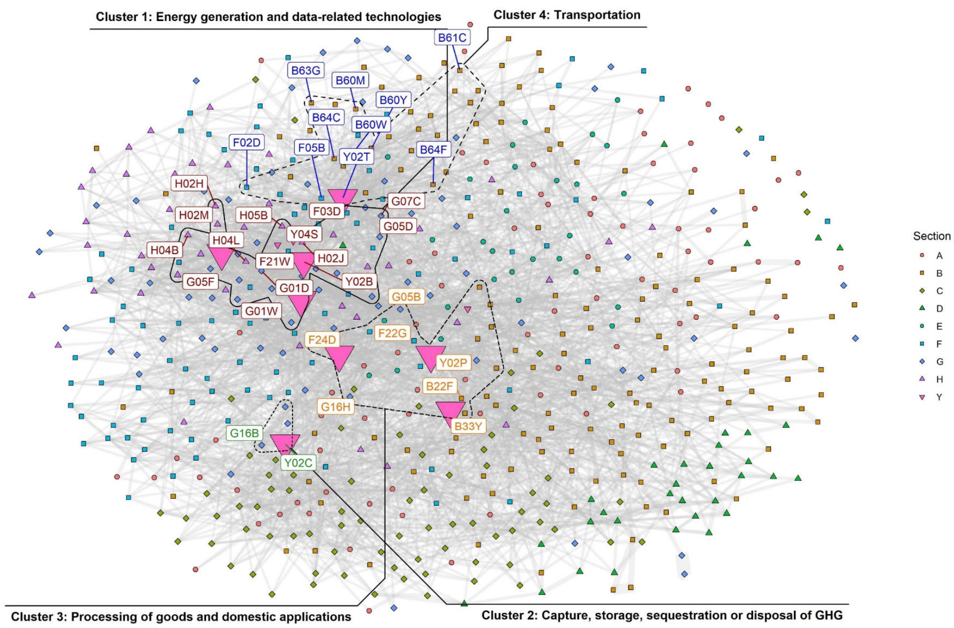
European and Japanese R&D Scoreboard companies have specialisations in at least one relevant technology of each cluster. Conversely, China and the US do not have any specialisations in the cluster related to processing of goods and domestic applications (Cluster 3) and to capture, storage, sequestration or disposal of GHG (Cluster 2) respectively. The EU leads in number of specialisations in transportation (Cluster 4) and in energy generation and data-related technologies (Cluster 1).

R&D Scoreboard companies from the US hold the same number of specialisations as the European Scoreboard companies in Cluster 1. Japanese R&D Scoreboard companies follow European ones in terms of specialisations in technologies related to processing of goods and domestic applications (Cluster 3). Other (non-R&D Scoreboard) actors seem to specialize more in technologies relevant for capture, storage, sequestration or disposal of GHG (Cluster 2), respectively, as well as related to processing of goods and domestic applications (Cluster 3).

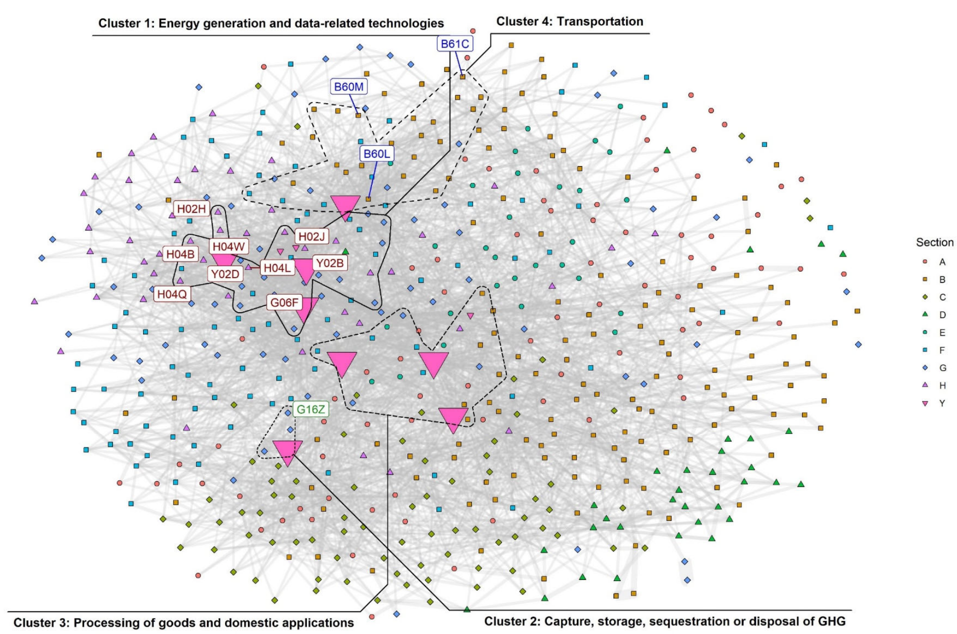
Now we proceed by further disaggregating specialisations of R&D Scoreboard companies linked to each cluster according to the four geographical regions by using the technology space approach (see Figure 12). We observe a stronger specialisation of European and Japanese R&D Scoreboard companies in sustainability technologies with existing specialisations in four Y02 codes. Both share specialisation in 'Climate change mitigation technologies related to transportation' (Y02T), 'Climate change mitigation technologies in the production or processing of goods' (Y02P), and 'Climate change mitigation technologies in the production or processing of goods' (Y02C). EU based R&D Scoreboard companies also show specialisations in the code 'Climate change mitigation technologies related to buildings' (Y02B), whereas Japanese ones possesses an exclusive specialisation in 'Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution' (Y02E). US-American and Chinese R&D Scoreboard companies, in turn, possess specialisations in only two Y02 codes. The R&D Scoreboard companies from the US specialise in 'Climate change mitigation technologies related to transportation' (Y02T) and 'Climate change mitigation technologies in ICTs' (Y02D). The latter specialisation applies also to Chinese R&D Scoreboard companies, which in addition have a specialisation advantage in 'Climate change mitigation technologies related to buildings' (Y02B). All four considered geographical areas miss specialisation advantages in two sustainable codes linked to Cluster 3, namely codes 'Technologies for adaptation to climate change' (Y02A) and 'sustainable technologies linked to wastewater treatment or waste management' (Y02W).

Figure 12: Technology space for EU (a), Chinese (b), Japanese (c), and US-American (d) R&D scoreboard companies (2016-2018).

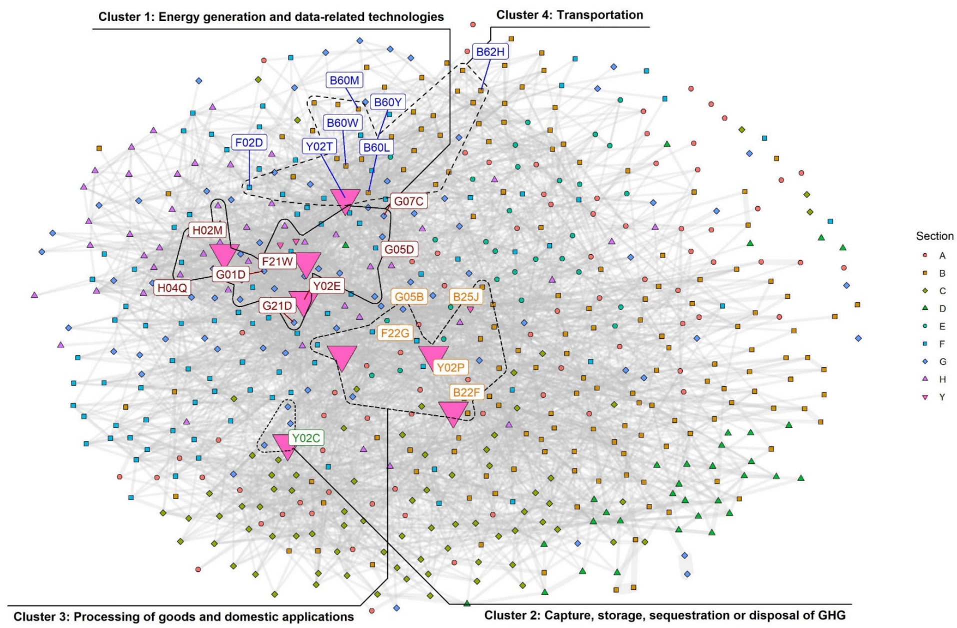
a)



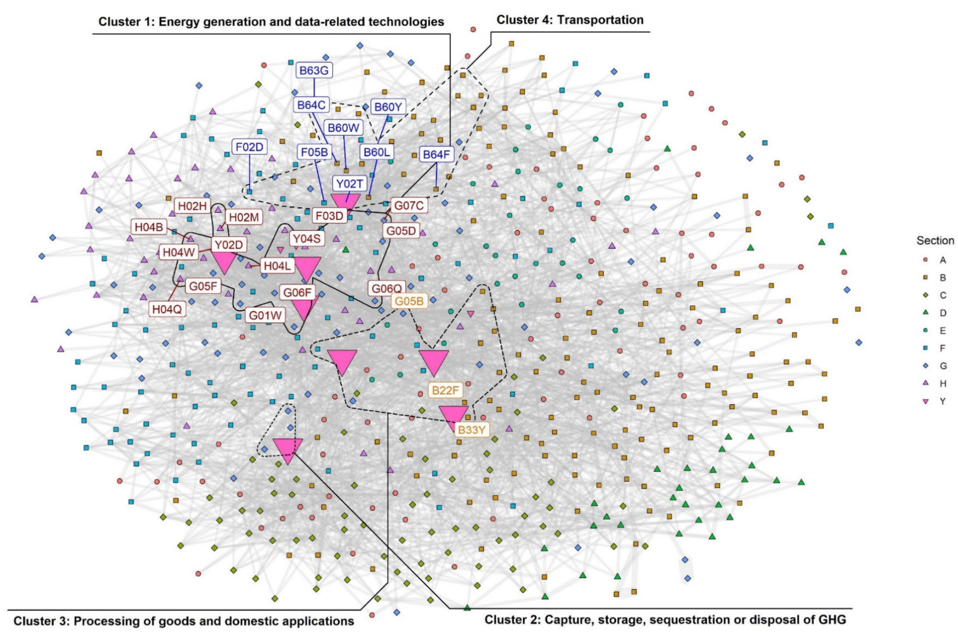
b)



c)



d)



Note: Specialisations based on the RTA index. The RTA used to calculate the specialisations compares the relative use of each technology by R&D Scoreboard companies from the respective countries in comparison to its global average use.
Source: PATSTAT 2019 for defining the technology space, and PATSTAT 2021 for calculating specialisations.

To summarize the results from the pervious technology space analysis, R&D Scoreboard companies from the US, China, Japan and the EU are well suited to develop sustainability inventions linked to transportation. Except for China, R&D Scoreboard companies from all geographic regions have a specialisation advantage in sustainable transportation technologies (Y02T) and share many specialisations in technologies related to Cluster 4. EU and Japanese R&D Scoreboard companies dominate many technologies relevant to the deployment of sustainable inventions in the area of processing of goods (Cluster 3). They share specialisations in the sustainable code linked to the production or processing of goods (Y02P) as well as in other relevant related technologies such as 'Working metallic powder' (B22F), 'Additive manufacturing' (B33Y), 'Healthcare informatics' (G16H), and 'Superheating of steam' (F22G). US-American and Chinese R&D Scoreboard companies, in turn, dominate many technologies relevant to the deployment of sustainable ICT inventions. Both regions are specialised in 'Climate change mitigation technologies in ICTs' (Y02D) as well as in related technologies such as 'Electrical digital data processing' (G06F), 'Data processing systems or methods (G06Q), 'Transmission of digital information' (H04L), and 'Wireless communication' (H04W).

3.2.2 Individual firm level

Next, we identify leading R&D Scoreboard companies in the relevant technologies identified previously as related to digital sustainability. We first analyse the specialisation advantages of the Top 60 R&D Scoreboard companies in these technologies. Then, we identify who are the top-performing R&D Scoreboard companies according to their number of relevant specialisations in each of the four identified clusters of relevant technologies. Conversely to the first set of Top 60 companies, which considers the total number of specialisations advantages in all possible relevant technologies, the second set of companies focuses on top-performing companies of each cluster based on their total number of relevant specialisations in the considered cluster.

Appendix E presents the Top 60 R&D Scoreboard companies identified for the first criterion (i.e., total number of specialisation advantages in relevant technologies related to digital sustainability). It is noted that although R&D Scoreboard companies have particularly large patent portfolios, some companies concentrate their specialisations in specific clusters (e.g., Panasonic and Toshiba in Cluster 1, Kawasaki Heavy Industries in Cluster 4) rather than distributing their specialisations more equally over all identified clusters. This pattern is even more pronounced for R&D Scoreboard companies with high shares of relevant specialisations for digital sustainability inventions. With the exception of Tesla, all R&D Scoreboard companies (e.g., Nari Technology, LS, Delta Electronics [Taiwan], Cypress Semiconductor, Schneider, Nio, Volkswagen, ABB, Rockwell Automation, Johnson Controls, and Telefonica) with a share equal to or above 30% of relevant specialisations focus on a particular cluster of related technologies.

Furthermore, in terms of the number of individual R&D Scoreboard companies with relevant specialisations, Japan leads (19), followed by the US (14), and South Korea, Germany and China (each with 5). The main Scoreboard sectors with most R&D Scoreboard companies with relevant specialisations are 'Automobiles & other transport' and 'ICT producers' (16 each), 'Industrials' (14), and 'ICT Services' (5). The leading ICB sectors, in turn, are 'Automobiles & Parts' (15), followed by 'Electronic & Electrical Equipment' (11), and 'General Industrials' (9).

Next, we focus on leaders from each identified relevant cluster. We select top-performing companies of each cluster based on their total number of relevant specialisations in the considered cluster. Starting with 'Energy generation and data-related technologies' (Cluster 1), we identify 31 leading R&D Scoreboard companies (see Table 9). The focus on 31 top-performing companies is an arbitrary choice, partially justified by the large number of leading companies (namely, 24 companies) "tied" in the following positions with 9 relevant specialisations. 21 out of the identified 31 leading companies from Cluster 1 are also in the list of Top 60 R&D Scoreboard companies (see Appendix E). The 10 remaining companies, compared to the companies presented in Appendix E, are Blackberry, ZTE, Infineon Technologies, Mediatek, Analog Devices, Motorola, On Semiconductor, Texas Instruments, Itron, and Intel. In terms of number of companies with relevant existing technological specialisations in Cluster 1, US-American R&D Scoreboard companies lead (9), followed by Japan (8), and China and Taiwan (3 each). The majority of R&D Scoreboard companies are in the Scoreboard sectors 'ICT producers' (21), followed by 'ICT services' (4). In terms of ICB sectors, 'Technology Hardware & Equipment' leads (14), followed by 'Electronic & Electrical Equipment' (7).

Table 9: R&D Scoreboard Companies with relevant specialisations in Cluster 1 (2016-2018).

Short Name	Country	Total Spec.	Irr. Spec.	Rel. Spec.	Share Rel. Spec.	Spec. Cl. 1	Spec. Cl. 2	Spec. Cl. 3	Spec. Cl. 4
Nxp Semiconductors	Netherlands	50	35	15	30%	15	0	0	0
Nari Technology	China	43	23	20	47%	15	1	3	1
Panasonic	Japan	141	117	24	17%	14	1	5	4
Hitachi	Japan	123	93	30	24%	13	2	7	8
Toshiba	Japan	104	84	20	19%	13	1	3	3
Nec	Japan	52	39	13	25%	13	0	0	0
Siemens	Germany	90	69	21	23%	12	1	6	2
Delta Electronics (Taiwan)	Taiwan	49	34	15	31%	12	0	1	2
Xiaomi	Cayman Islands	48	35	13	27%	12	0	1	0
Microchip Technology	US	33	21	12	36%	12	0	0	0
Cypress Semiconductor	US	28	13	15	54%	12	0	3	0
Mitsubishi Electric	Japan	69	51	18	26%	11	0	3	4
Thales	France	54	42	12	22%	11	1	0	0
Inventec	Taiwan	54	39	15	28%	11	0	3	1
Analog Devices	US	41	30	11	27%	11	0	0	0
Schneider	France	31	17	14	45%	11	0	3	0
Telefonica	Spain	15	3	12	80%	11	0	1	0
Honeywell	US	114	92	22	19%	10	1	7	4
Omron	Japan	71	55	16	23%	10	0	4	2
Hisense Electric	China	72	56	16	22%	10	1	3	2
Denso	Japan	96	81	15	16%	10	0	0	5
NTT	Japan	46	34	12	26%	10	1	0	1
Blackberry	Canada	35	24	11	31%	10	0	0	1
ZTE	China	24	13	11	46%	10	1	0	0
Infineon Technologies	Germany	43	33	10	23%	10	0	0	0
On Semiconductor	US	43	33	10	23%	10	0	0	0
Texas Instruments	US	41	31	10	24%	10	0	0	0
Intel	US	36	26	10	28%	10	0	0	0
Mediatek	Taiwan	32	22	10	31%	10	0	0	0
Motorola	US	30	20	10	33%	10	0	0	0
Itron	US	15	5	10	67%	10	0	0	0

Source: PATSTAT 2021b.

We know that Cluster 2 with 'technologies related to the capture, storage, sequestration or disposal of GHGs' is relatively small. We find 58 R&D Scoreboard companies that present at least one specialisation in related technologies from this cluster. We identify the Top 6 R&D Scoreboard companies of these (see Table 10). Once again, the selection of top-performing companies is an arbitrary choice made based on the number of leading companies "tied" in the following positions (namely, there are 52 companies with at least 1 specialisation in this cluster). From this set of 6 leaders, Fujitsu and IBM are not part of the Top 60 list. In this cluster leaders, we find three R&D Scoreboard companies from the US, two from Japan, and one from China. Three of them are in Scoreboard sector 'ICT services' (all from ICB sector 'Software & Computer Services'), two from 'Industrials' (ICB sector 'General Industrials'), and one from 'ICT producers' (ICB sector 'Electronic & Electrical Equipment').

Table 10: R&D Scoreboard Companies with relevant specialisations in Cluster 2 (2016-2018).

Short Name	Country	Total Spec.	Irr. Spec.	Rel. Spec.	Share Rel. Spec.	Spec. Cl. 1	Spec. Cl. 2	Spec. Cl. 3	Spec. Cl. 4
Hitachi	Japan	123	93	30	24%	13	2	7	8
General Electric	US	143	119	24	17%	9	2	7	6
Alphabet	US	60	45	15	25%	9	2	3	1
Fujitsu	Japan	55	45	10	18%	7	2	1	0
IBM	US	35	25	10	29%	7	2	1	0
Metallurgical of China	China	98	84	14	14%	6	2	6	0

Source: PATSTAT 2021b.

We find 26 R&D Scoreboard companies with specialisation advantages in 'technologies related to the processing of goods and domestic applications' (Cluster 3) (see Table 11). Once again, the selection of top-performing companies is an arbitrary choice made based on the number of leading companies "tied" in the following positions (namely, there are 17 companies with at least 4 relevant specialisations in this cluster). 15 of the leading R&D Scoreboard companies from this third cluster are also amongst the Top 60 performers in terms of total number of relevant specialisations. The remaining 11 additional companies that are not part of the 60 leaders list are: Schlumberger, Linde, Weatherford International, Daiichi Sankyo, Nippon Steel, Seiko Epson, Jtekt, Sandvik, Kinpo Electronics, Harris, and Lincoln Electric. In terms of the number of R&D companies leading in this cluster, Japan leads (9), followed by US (8), and Germany (3). The majority of R&D Scoreboard companies are in the Scoreboard sectors 'Industrials' (9) and 'ICT producers' (6). In terms of ICB sectors, we find with 'General Industrials' (4) in the lead, followed by 'Aerospace & Defence', 'Automobiles & Parts', 'Electronic & Electrical Equipment', 'Industrial Engineering', and 'Technology Hardware & Equipment' (3 each).

Table 11: R&D Scoreboard Companies with relevant specialisations in Cluster 3 (2016-2018).

Short Name	Country	Total Spec.	Irr. Spec.	Rel. Spec.	Share Rel. Spec.	Spec. Cl. 1	Spec. Cl. 2	Spec. Cl. 3	Spec. Cl. 4
Hitachi	Japan	123	93	30	24%	13	2	7	8
General Electric	US	143	119	24	17%	9	2	7	6
Honeywell	US	114	92	22	19%	10	1	7	4
Metallurgical of China	China	98	84	14	14%	6	2	6	0
Siemens	Germany	90	69	21	23%	12	1	6	2
Mitsubishi Heavy	Japan	147	127	20	14%	8	1	6	5
Mitsubishi Motors	Japan	113	92	21	19%	6	1	6	8
Kobe Steel	Japan	115	103	12	10%	4	1	6	1
Emerson Electric	US	83	68	15	18%	9	0	6	0
Boeing	US	132	115	17	13%	7	0	6	4
Daiichi Sankyo	Japan	81	71	10	12%	4	0	6	0
Panasonic	Japan	141	117	24	17%	14	1	5	4
Bae Systems	UK	90	73	17	19%	6	0	5	6
United Technologies	US	131	114	17	13%	6	1	5	5
Robert Bosch	Germany	120	104	16	13%	6	0	5	5
Tesla	US	32	18	14	44%	6	0	5	3
Jtekt	Japan	55	44	11	20%	4	0	5	2
Harris	US	31	21	10	32%	4	0	5	1
Weatherford International	Ireland	65	55	10	15%	2	1	5	2
Seiko Epson	Japan	74	66	8	11%	2	1	5	0
Linde	Germany	49	42	7	14%	1	1	5	0
Nippon Steel	Japan	63	57	6	10%	1	0	5	0
Schlumberger	Curazao	44	38	6	14%	1	0	5	0
Kinpo Electronics	Taiwan	20	14	6	30%	1	0	5	0
Lincoln Electric	US	20	14	6	30%	1	0	5	0
Sandvik	Sweden	35	26	9	26%	0	0	5	4

Source: PATSTAT 2021b.

Finally, we highlight 14 R&D Scoreboard companies with specialisation advantages in 'technologies related to transportation' (Cluster 4) (see Table 12). Another 25 companies, not shown in Table 12, present at least 5 specialisations in technologies relevant to Cluster 4. 10 of the identified 14 leading companies from this cluster are also among the top 60 performers regarding total number of relevant specialisations. The additional R&D Scoreboard companies that are not part of the 60 leaders list are: Yamaha Motor, Yamaha, Subaru, and Textron. The majority of companies comes from Japan (9), followed by the US (3), and the UK and China (1 each). In terms of Scoreboard sector, 'Automobiles & other transport' lead (5 companies, all of them from the ICB sector 'Automobiles & Parts'), followed by 'ICT producers' (3 companies, all from 'Electronic & Electrical Equipment') and 'Industrials' (also 3 companies in total, 2 being from the ICB sector 'General Industrials' and 1 from 'Industrial Engineering').

Table 12: R&D Scoreboard Companies with relevant specialisations in Cluster 4 (2016-2018).

Short Name	Country	Total Spec.	Irr. Spec.	Rel. Spec.	Share Rel. Spec.	Spec. Cl. 1	Spec. Cl. 2	Spec. Cl. 3	Spec. Cl. 4
Kawasaki Heavy Industries	Japan	109	94	15	14%	1	1	2	11
Hitachi	Japan	123	93	30	24%	13	2	7	8
Mitsubishi Motors	Japan	113	92	21	19%	6	1	6	8
Byd	China	68	53	15	22%	7	0	1	7
Honda Motor	Japan	88	75	13	15%	5	1	0	7
Subaru	Japan	70	59	11	16%	4	0	0	7
Yamaha Motor	Japan	43	35	8	19%	1	0	0	7
General Electric	US	143	119	24	17%	9	2	7	6
Bae Systems	UK	90	73	17	19%	6	0	5	6
Ntn	Japan	60	43	17	28%	8	0	3	6
Ford Motor	US	91	75	16	18%	9	0	1	6
Textron	US	49	39	10	20%	3	0	1	6
Sumitomo Electric	Japan	106	91	15	14%	8	1	0	6
Yamaha	Japan	66	59	7	11%	1	0	0	6

Source: PATSTAT 2021b.

Overall, there are 66 unique R&D Scoreboard companies when considering the existing technological specialisation for each of the four Clusters of relevant technologies for digital sustainability inventions. The majority comes from Japan (23), followed by the US (20), China (5), Germany and Taiwan (4 each). Aggregated at EU27 level, there are 11 EU companies among the identified R&D Scoreboard companies. R&D Scoreboard companies from the US dominate in 'Energy generation and data-related technologies' (Cluster 1) and 'technologies related to the capture, storage, sequestration or disposal of GHG' (Cluster 2), whereas Japanese ones prevail in 'technologies related to the processing of goods and domestic applications' (Cluster 3), and 'technologies related to transportation' (Cluster 4). The Scoreboard sectors of these R&D Scoreboard companies are 'ICT producers' (27), 'Industrials' (8), and 'ICT services' (7), whereas the top-performing ICB sectors are 'Technology Hardware & Equipment' (17), 'Electronic & Electrical Equipment' (10), and 'Automobiles & Parts' (8). Each cluster has a distinct prevailing Scoreboard sector: 'ICT producers' are the major leaders in the 'energy generation and data-related technologies' (Cluster 1), 'ICT services' in 'technologies related to the capture, storage, sequestration or disposal of GHG' (Cluster 2), 'Industrials' in the 'processing of goods and domestic applications' cluster (Cluster 3), and 'Automobiles & other transport' in the 'technologies related to transportation' cluster (Cluster 4).

3.2.3 European R&D Scoreboard companies

From an EU perspective, technological specialisation advantages exist in 'Energy generation and data-related technologies' (Cluster 1) as well as 'technologies related to transportation' (Cluster 4). When we take a closer look at the 'existing' and 'missing' specialisation advantages of European R&D scoreboard companies in the relevant technologies of Cluster 1 (see Table 13), we observe that European R&D scoreboard companies specialise in technologies related to wind power (F03D), but miss specialisations in other alternative sources of energy, like solar and nuclear power-related technologies (H02S and G21D, respectively). They master most technologies related to the transmission, distribution and management of electric power systems (i.e., F21W, G05F, H02H, H02J, H02M, H04B, H05B, Y02B, Y04S), whilst lacking only a few of related specialisations (i.e., G08C, Y02E). In view of data-related technologies, European R&D scoreboard companies hold only one specialisation advantage in 'Transmission of digital information' (H04L), while lacking many other conducive specialisations such as 'Electric digital data processing' (G06F), 'Data processing systems or methods' (G06Q), 'Wireless communication networks' (H04W), and 'Climate change mitigation technologies in ICTs' (Y02D).

Table 13: Available and missing specialisations of European scoreboard companies in the relevant technologies of Cluster 1.

CPC	Description	Specialis.
F03D	Wind motors	Available
F21W	Indexing scheme associated with subclasses F21K, F21L, F21S and F21V, relating to uses or applications of lighting devices or systems	Available
G01D	Measuring not specially adapted for a specific variable; arrangements for measuring two or more variables not covered in a single other subclass; tariff metering apparatus; measuring or testing not otherwise provided for	Available
G01W	Meteorology	Available
G05D	Systems for controlling or regulating non-electric variables	Available
G05F	Systems for regulating electric or magnetic variables	Available
G06F	Electric digital data processing	Missing
G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for	Missing
G07C	Time or attendance registers; registering or indicating the working of machines; generating random numbers; voting or lottery apparatus; arrangements, systems or apparatus for checking not provided for elsewhere	Available
G08C	Transmission systems for measured values, control or similar signals	Missing
G21D	Nuclear power plant	Missing
H02H	Emergency protective circuit arrangements	Available
H02J	Circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy	Available
H02M	Apparatus for conversion between ac and ac, between ac and dc, or between dc and dc, and for use with mains or similar power supply systems; conversion of dc or ac input power into surge output power; control or regulation thereof	Available
H02S	Generation of electric power by conversion of infra-red radiation, visible light or ultraviolet light, e.g. Using photovoltaic [PV] modules	Missing
H04B	Transmission	Available
H04L	Transmission of digital information, e.g. Telegraphic communication	Available
H04Q	Selecting	Missing
H04W	Wireless communication networks	Missing
H05B	Electric heating; electric light sources not otherwise provided for; circuit arrangements for electric light sources, in general	Available
Y02B	Climate change mitigation technologies related to buildings, e.g. Housing, house appliances or related end-user applications	Available
Y02D	Climate change mitigation technologies in information and communication technologies [ICT], i.e. Information and communication technologies aiming at the reduction of their own energy use	Missing
Y02E	Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution	Missing
Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. Smart grids	Available

Source: Authors.

Regarding 'technologies related to transportation' (Cluster 4) most of the relevant specialisation advantages are present (see Table 14). However, European R&D scoreboard companies lack a specialisation advantage in 'technology related to electrical vehicles' (B60L). This particular technology is the second largest of this cluster (as shown previously in Table 6), besides also being one of the fastest growing and arguably important technology to the deployment of sustainable transportation-related inventions.

Table 14: Available and missing specialisations of European scoreboard companies in the relevant technologies of Cluster 4.

CPC	Description	Specialis.
B60L	Propulsion of electrically-propelled vehicles; supplying electric power for auxiliary equipment of electrically-propelled vehicles; electrodynamic brake systems for vehicles in general; magnetic suspension or levitation for vehicles; monitoring operating variables of electrically-propelled vehicles; electric safety devices for electrically-propelled vehicles	Missing
B60M	Power supply lines, and devices along rails, for electrically-propelled vehicles	Available
B60W	Conjoint control of vehicle sub-units of different type or different function; control systems specially adapted for hybrid vehicles; road vehicle drive control systems for purposes not related to the control of a particular sub-unit	Available
B60Y	Indexing scheme relating to aspects cross-cutting vehicle technology	Available
B61C	Locomotives; motor railcars	Available
B62H	Cycle stands; supports or holders for parking or storing cycles; appliances preventing or indicating unauthorized use or theft of cycles; locks integral with cycles; devices for learning to ride cycles	Missing
B63G	Offensive or defensive arrangements on vessels; mine-laying; mine-sweeping; submarines; aircraft carriers	Available
B64C	Aeroplanes; helicopters	Available
B64F	Ground or aircraft-carrier-deck installations specially adapted for use in connection with aircraft; designing, manufacturing, assembling, cleaning, maintaining or repairing aircraft, not otherwise provided for; handling, transporting, testing or inspecting aircraft components, not otherwise provided for	Available
F02D	Controlling combustion engines	Available
F03B	Machines or engines for liquids	Missing
F05B	Indexing scheme relating to wind, spring, weight, inertia or like motors, to machines or engines for liquids covered by subclasses F03B, F03D and F03G	Available
Y02T	Climate change mitigation technologies related to transportation	Available

Source: Authors.

4 SUMMARY OF MAIN FINDINGS

This report offered a novel approach to identify patents associated with digital sustainability technologies, which is complementary to extant identification strategies (see for example, Amoroso et al. 2921). We developed an identification strategy based upon six search modules, which combine specialists' opinion, keywords and classification-based approaches and we used the 'Y' section of the CPC classification as our main reference (USPTO, 2021). The reported accuracy of the strategy is 95%. Given that our identification of 'digital' components included ICT-technologies but is not exclusively restricted to an ICT-classification, we search in a 'broader way'. This approach might allow us to capture, for example, not only ICT prominently in 'Energy generation and data-related technologies' but also in 'Technologies related to the 'Processing of goods and domestic applications'. Given the dominance of US and Chinese R&D scoreboard companies in ICT-related technologies, our identification might also capture other relevant specializations related to digital sustainability inventions by European R&D Scoreboard firms However, this remains subject to further research.

4.1 Evolution of digital sustainability technologies

Based on a technology space approach, we contributed to the understanding of the nature and evolution of digital sustainability technologies globally. This approach allowed us to show how technological specialisations related to digital sustainability patents emerged and evolved in the period from 2001 to 2018. We detected four main clusters of technologies used to create digital sustainability inventions, which rely upon 53 relevant technologies. The four clusters occupy distinct areas of a technology space, which indicates that technology similarity within these clusters is higher than outside them. The identified clusters are 'Energy generation and data-related technologies' (Cluster 1); 'Technologies related to the capture, storage, sequestration or disposal of GHGs' (Cluster 2); 'Technologies related to the processing of goods and domestic applications (Cluster 3); and 'Technologies related to transportation' (Cluster 4).

In terms of scope, 'Energy generation and data-related technologies' (Cluster 1) and 'Technologies related to the processing of goods and domestic applications' (Cluster 3) have the largest number of relevant related technologies throughout the observation period. 'Technologies related to the processing of goods and domestic applications' (Cluster 3) and 'Technologies related to transportation' (Cluster 4) increased in the number of specialised technologies. In fact, 'Technologies related to the processing of goods and domestic applications' (Cluster 3) is now the second largest cluster and the only that continuously grows over time in terms of the number of specialisations. In contrast, 'Technologies related to the capture, storage, sequestration or disposal of GHGs' (Cluster 2) is the smallest cluster and lost all specialised technologies towards the end of the observation period (2001-2018). This evidence indicates very different dynamics in the evolution cluster of related technologies linked to digital sustainability inventions.

4.2 Role of R&D Scoreboard companies

The general insights about the nature and evolution of digital sustainability technologies provided an appropriate background for the analysis of the role of the world's Top 2,000 corporate R&D investors. We performed further analysis using data from the JRC-OECD COR&DIP© v.3 dataset (2021 version). When analysing the digital sustainability technologies created by EU Industrial R&D Scoreboard companies, we focused on the period from 2016 to 2018. We updated the initial dataset by applying our identification strategy to the PATSTAT 2021b version. In this part of the analysis, we considered only IP5 patents.

For 2016 to 2018, 7,808 priorities are associated with digital sustainability technologies, from which scoreboard companies own 5,057. While R&D scoreboard companies are responsible for filing 60.6% of all IP5 patents globally, this proportion reaches 64.8% for digital sustainability IP5 patents. This concentration underlines the importance of R&D Scoreboard companies in technology development and diffusion in the era of the twin transition. However, for the observation period, only 1.55% of all priority patents by R&D Scoreboard companies, and 14.11% of all their sustainable priority patents, are digital sustainability inventions. This indicates that the integration of the 'digital' and 'green' transition is still in a very early phase.

In terms of geography, we find 25 out of 40 countries linked to R&D scoreboard companies with at least one digital sustainability patent during the observation period. The Top 5 countries are the US, Japan, China, South Korea and

Germany. In case of the US and Chinese companies, their proportional contribution to digital sustainability technologies exceeds their respective contributions to general patenting. Amongst the Top 15 countries, we find also, apart from Germany, other European locations such as Ireland, France, Netherlands, Sweden, UK and Switzerland. In case of Ireland and Sweden, their proportional contribution to digital sustainability technologies exceeds their respective contributions to general patent application by R&D scoreboard companies.

In terms of sectoral distribution, we find that R&D scoreboard companies from the Scoreboard sectors 'ICT producers', 'Industrials', and 'Automobile and other transports' produce most digital sustainability patents. However, the vast majority is concentrated amongst 'ICT producers'. A lower level of disaggregation documents the importance of two Scoreboard sectors: 'Technology Hardware & Equipment' and 'Electronic & Electrical Equipment', whereby the former contributes proportionally more in comparison to general patent output of scoreboard companies. Thus, it seems that the core of digital sustainability inventions originates from digital-related companies embracing sustainable technologies.

Only about 20% of the 2,000 R&D scoreboard companies have at least one digital sustainability patent registered between 2016 and 2018. The Top 50 and Top 10 scoreboard companies account for 66.2% and 32.3% respectively from all digital sustainability inventions identified. The Top-3 Scoreboard companies in terms of digital sustainability patents output are Qualcomm, Intel, and General Electric, which are all located in the US. Amongst the remaining Top-10, we find Samsung Electronics (South Korea), Fanuc (Japan), Siemens (Germany), Ford Motor (US), Hitachi (Japan), Zte (China) and Huawei (China). Amongst, the Top-50 R&D scoreboard companies in terms of digital sustainable output, we find only four from the EU: Siemens (Germany), BMW (Germany), Schneider (France), and Johnson Controls (Ireland).

4.3 Technological specialisation of European R&D Scoreboard companies

In terms of technological capabilities amongst R&D scoreboard companies, the European R&D scoreboard companies lead in number of specialisations in technologies relevant to the deployment of digital sustainability inventions, followed by the US, Japan, and China. A specialisation in these relevant technologies indicates the presence of relevant knowledge needed to create digital sustainability inventions, even though digital sustainability patents might not yet materialise. In particular, European scoreboard companies lead in number of specialisations in transportation (Cluster 4) and in energy generation and data-related technologies (Cluster 1). The US companies are equally specialised as their European peers in Cluster 1. Japanese companies follow European ones regarding number of specialisations held in technologies related to the processing of goods and domestic applications (Cluster 3). Compared to the rest of the world, European, US-American, and Japanese companies are particularly specialised in technologies relevant to the deployment of inventions linked to transportation (Cluster 4). Considering the main differences, European and Japanese companies dominate many technologies relevant to the deployment of digital sustainability inventions in the area of 'processing of goods and domestic applications' (Cluster 3). US-American and Chinese companies, in turn, dominate technologies relevant to the deployment of sustainable ICT inventions.

Finally, we investigated the specialisation of R&D scoreboard companies in relevant technologies of the identified clusters. Although R&D scoreboard companies own particularly large patent portfolios, we find that these companies concentrate their specialisations in specific clusters of digital sustainability technologies. We find that R&D scoreboard companies concentrate most of their inventions in one specific cluster. Therefore, the leaders of each cluster are distinct: 85.7% (66 out of 77) of the identified R&D scoreboard companies in each cluster appear as leaders in just a single cluster.

The majority of these Top 66 companies originate from Japan, followed by the US, China, Germany and Taiwan. At the EU27 level, there are 11 European scoreboard companies. US-based scoreboard companies prevail in 'Energy generation and data-related technologies' (Cluster 1) and 'Technologies related to the capture, storage, sequestration or disposal of greenhouse gases' (Cluster 2), whereas Japanese ones dominate in 'Technologies related to the processing of goods and domestic applications' (Cluster 3) and 'Technologies related to transportation' (Cluster 4). R&D scoreboard companies based in Europe contribute primarily in clusters 3 and 4.

Despite the fact that European companies lead in number of specialisations in technologies relevant to the deployment of digital sustainability inventions, their generation of digital sustainability inventions lags behind R&D scoreboard companies from the US and Japan. We link this to a lack of particular technological specialisations relevant for the specific clusters identified. For the largest cluster of 'Energy generation and data-related technologies' (Cluster 1), European R&D companies miss specialisations in other alternative sources of energy, like solar and nuclear power-related technologies. They also lack relevant specialisations in 'Transmission systems for measured values, control or similar signals' as well as 'Reduction of GHG emissions related to energy generation, transmission or distribution'. In respect to data-related technologies of Cluster 1, European scoreboard companies lack most of the relevant specialisations such as 'Electric digital data processing', 'Data processing systems or methods', 'Wireless communication networks', and 'CCMTs in ICT'. In respect to digital sustainability inventions related to transportation (Cluster 4), European R&D scoreboard companies possess most of the relevant specialisations, apart from capabilities related to electrical vehicles, which is the second largest and one of the fastest-growing technologies in this cluster.

5 POLICY IMPLICATIONS

The application of a technology space perspective is particularly suitable for the development of the so-called 'Smart policies'. These policies focus on leveraging existing specialisation and capabilities of sub-national regions and countries as a means to strengthen them and to allow for the emergence of new competences and knowledge in related technologies (Foray et al., 2009; Foray et al., 2011). Such policies have been increasingly adopted by many countries, e.g., in Europe (through its Smart Specialization Strategy), China, Mexico, Canada, etc. (Hidalgo, 2021). In the EU, the focus varies from a broader smart policy strategy (e.g., Balland et al. 2019), to a narrower one targeting the development of specific technologies (see Montresor and Quatraro (2020). In the latter case, for example, it is found that the development of capabilities in key enabling technologies facilitates the transition toward more sustainable technological trajectories (Montresor and Quatraro, 2020).

From this point of view, a smart policy perspective for digital sustainability technologies would focus on leveraging existing specialisations in similar relevant technologies, which we identified in this report by clusters based on a technology space approach. Given the higher similarity of technologies within the identified clusters, smart policies should focus on further developing competencies from the clusters where countries already have some existing technological advantages. From a European perspective, specialisations exist in technologies relevant for digital sustainability inventions in the area of 'Energy generation and data-related technologies' (Cluster 1) and 'Technologies related to transportation' (Cluster 4).

Our analysis also indicated a lack of various related specialisation in the respective clusters. Therefore, a European smart policy strategy could focus, for example, on leveraging the further development of technological specialisations linked to alternative sources of electric power generation and to the transmission/distribution and management of electric power systems to improve the output in digital sustainability inventions in the area of energy generation and data-related technologies. The fact that these competencies are related to the ones needed to deploy data-related technologies suggests that competencies linked to the latter can emerge through related diversification, which would allow for better coupling with missing digital components, in particular, in 'Electric digital data processing', 'Data processing systems or methods', 'Wireless communication networks' and 'CCMTs in ICT'.

Moreover, the further development of specialisations related to digital sustainability inventions in transportation technologies could favour the emergence of new capabilities needed for the deployment of electric vehicle-related inventions. Conversely to the technologies from Cluster 1, where many European corporate leaders were identified (e.g., Siemens, Robert Bosch, NXP Semiconductors, Eaton Corporation, Schneider, ABB, Johnson Controls, Thales, Telefonica, and Infineon Technologies), there are no European leaders (except for BMW) identified in Cluster 4 (transportation). This indicates that, despite having large number of leading companies in the 'Automobiles & Parts' scoreboard sector, these leaders still fall behind Japanese and US peers in the creation of digital sustainability inventions in the transportation sector.

The 'twin transition' refers to the potential of digital technologies to enable a more sustainable future through increasing energy and resource efficiency (Amoroso et al., 2021). The findings in this report suggest that, for digital sustainability technologies, the core of inventions comes from digital-related companies embracing sustainable technologies rather than sustainability-related companies embracing digital-related technologies. The vast majority of leaders in these technologies are 'ICT producers' from ICB sectors related to 'Technology Hardware & Equipment' and 'Electronic & Electrical Equipment'. This corroborates Amoroso et al. (2021), who also find a high proportion of CCMT-related patents being related to ICT industries.

In sum, our findings suggest that large and established corporate R&D investors advance the 'twin transition' by adding sustainability components to their existing technologies. We can confirm that the majority of digital sustainability inventions comes from the top R&D investors, which underlines their importance for technology diffusion in the era of the twin transition. However, we still are in a very early phase of the integration of the 'digital' and 'green' transition. This actually strengthens the argument for the use of smart policies to trigger the development of digital sustainability inventions through the leveraging of existing specialisations. This aligns with Montresor and Quatraro (2020), who highlight that related non-green knowledge can be even more important than green knowledge in the emergence of green technologies. Thus, the 'twin transition' might advance through the combination of sustainability components into the ‘regular’ inventions created by traditional and large corporate R&D investors. In this context, ICT sectors play a key role not only because they are the main developers of digital sustainability inventions, but also because they traditionally provide technologies to other sectors (Amoroso et al., 2021). This crucial role of the ICT sector, however, makes the leadership of European R&D scoreboard companies in digital sustainability technologies rather challenging, given the lack of globally competitive actors in ICT sectors such as ‘Technology Hardware & Equipment’.

Part II: Methodological Note

In this part of the report, the specialists present details on the identification strategy (Section 1), analysis and validation of the proposed strategy (Section 2), and the adjustments made to the patent dataset for making it more comparable to the EU climate neutrality report 2021 edition (Section 3), and information on access to data and code (Section 4). In this report, we use two distinct patent datasets. The first identifies priority patents based on PATSTAT's DOCDB simple family identifier (i.e., 'docdb_family_id') from the PATSTAT 2019a edition. We use the first dataset as a basis for the analysis presented in Section 2 of Part I of this report. The second dataset uses the IP5 strategy presented in Dernis et al. (2015) to identify priority patents using the PATSTAT 2021b edition. We uses the second dataset for analyses of R&D scoreboard activities presented in Section 3 of Part I in this report. Part II of this report offer a corresponding methodological note. In particular, we present details on the identification strategy of digital sustainability inventions (Section 1), validate the search outcomes (Section 2), provide information on the adjustments to the initial dataset for the analysis at the firm level (Section 3), and offer information on access to data and code (Section 4).

1 METHODS APPLIED IN THE IDENTIFICATION STRATEGY

All principal identification strategies have their limitations: Keyword-based strategies suffer from biases generated by the 'trending' of distinct terms and keywords over time. Classification-based strategies suffer from a lack of specificity or clear delimitation between distinct technologies, once classification codes are used with the intent of broadly highlighting technical mechanisms instead of intentionally distinguishing boundaries between them. Opinion-based approaches, in turn, are biased by the subjectivity generated through the particular perception of the experts. Rather than considering only if patents match a pre-selected list of keywords or classification codes, specialists read additional information about each patent to better understand their content and applicability, which puts an additional weight on the specialist's subjective judgement. Arguably, the individual limitations of these three main strategies are amplified when one uses them to identify emerging technologies that are highly complex and combine distinct technical components, as in the case of digital sustainability technologies.

Table 15: Search modules for the identification of digital sustainability inventions.

Module	Description of identification	Identification strategy
1	Patents tagged with both Y02 and Y04 codes	Specialists' opinion
2	Patents with at least one digital and one sustainability keyword in their title or abstract	Keyword-based
3	Patents with at least one digital keyword in the title or abstract and that are also classified under the Y02 code	Specialists' opinion & keyword-based
4	Patents that have at least one AI-related keywords in the title or abstract and are also classified under the Y02 code	Specialists' opinion & keyword-based
5	Patents classified with at least one of the considered digital-related IPC groups and also classified under the Y02 code	Classification-based & specialists' opinion
6	Patents classified with at least one of the considered digital-related IPC subclasses and also classified under the Y02 code	Classification-based & specialists' opinion

Source: Authors.

Considering the limitations of different individual identification strategies and the emerging character of digital sustainability technologies, we propose a search that combines specialists' opinions, keywords, and classification codes to mitigate their individual limitations. We use the selected strategies to create six distinct search modules, which can be added individually to create a joint comprehensive dataset of patent applications related to digital sustainability technologies (see Table 15 for an overview). Module 1 is entirely based on the Y section of the Cooperative Patent Classification (CPC). Module 2 is based only on keywords. Modules 3 and 4 combine the use of keywords to proxy digital and AI-related technologies with the Y02 code as a proxy for sustainable technologies. Modules 5 and 6 apply International Patent Classification (IPC) codes identified as typical for digital technologies to collect patents classified under the Y02 sustainability tag

1.1 Specialists' opinion

The 'Y' section in the CPC classification was launched in 2013 by a joint effort between the EPO and USPTO, combining algorithm-based identification with specialists' opinions (Angelucci et al., 2018). There are alternatives to this 'Y' scheme that also aim at identifying green technologies such as the OECD ENV-TECH classification (see Hascic and Migotto, 2015)6. The OECD ENV-TECH classification uses exclusively classification codes as a reference to identify green technologies. Therefore, the 'Y' scheme offers an additional quality filter by using specialists' opinion, which lowers false positives compared to the OECD ENV-TECH classification. In fact, the reported error rate for the 'Y' scheme is less than 7% (Angelucci et al. 2018). For a comparison between the OECD ENV-TECH classification and the 'Y' scheme, please see Section 2.2 below.

The 'Y' section in the CPC classification identifies both green (code Y02) and promising digital (code Y04) technologies. In this case, 'green' refers to technologies or applications for mitigation or adaptation against climate change. 'Promising digital', in turn, refers to information or communication technologies (ICT) having an impact on other

[6] https://www.oecd.org/environment/consumption-innovation/ENV-tech%20search%20strategies,%20version%20for%20OECDstat%20(2016).pdf

technology areas. Both codes also have subclasses allowing a finer-grained analysis of the respective technologies (see Table 16). There are more differentiated codes available within the Y02 'green' code, whereas the Y04 digital one contains only a single one related primarily to digital technologies in electric power. The broader scope of Y02 over Y04 is also reflected in the number of patents identified in each of them. Using PATSTAT 2019a, we find 3,931,281 sustainability patents identified through code Y02, and 63,120 digital patents identified through code Y04.

As digital sustainability technologies develop in the overlap between green and digital, the specific focus of the Y04 code limits the coverage of digital technologies and which introduces a bias in its overlap with green technologies towards digital technologies related to electric power. Therefore, it seems appropriate to broaden the search for digital technologies beyond the Y04 classification.

This seems appropriate given the widespread use and impact of digital technologies, which are dispersed and integrated into a variety of sectors beyond the electric one considered in the Y04 classification. Nevertheless, the Y classification is very accurate and seems the right choice as the main reference for the proposed search approach in this report, given that the Y02 code covers widely sustainability-related technologies. Thus, in our identification strategy five out of six search modules adopt the Y02 code as a reference for identifying sustainable technologies, whereby four modules aim at increasing the coverage of digital sustainability technologies through broadening the identification of digital technologies.

Table 16: Y classification for codes Y02 and Y04 and their respective descriptions.

Code	Sub-classification	Description
Y02		**Technologies or applications for mitigation or adaptation against climate change.**
	Y02A	Technologies for adaptation to climate change.
	Y02B	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications.
	Y02C	Capture, storage, sequestration or disposal of greenhouse gases [GHG].
	Y02D	Climate change mitigation technologies in information and communication technologies [ICT], i.e. information and communication technologies aiming at the reduction of their own energy use.
	Y02E	Reduction of greenhouse gas [GHG] emissions, related to energy generation, transmission or distribution.
	Y02P	Climate change mitigation technologies in the production or processing of goods.
	Y02T	Climate change mitigation technologies related to transportation.
	Y02W	Climate change mitigation technologies related to wastewater treatment or waste management.
Y04		**Information or communication technologies having an impact on other technology areas.**
	Y04S	Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grids.

*Source***:** USPTO (2021).

Alternative identification strategies (see Amoroso et al., 2021) relied exclusively on an ICT classification (Inaba and Squicciarini, 2017) to capture the digital component in the identification of digital sustainability patents. In comparison, our strategy aims to include ICT-technologies without being exclusively restricted to an ICT classification. Therefore, we combine keyword-based strategies with classification-based strategies to capture digital components in digital sustainability inventions more broadly.

1.2 Keyword-based search strategy

Given the limitation of the Y04 code, the keyword-based strategy is particularly important to extend the coverage of digital technologies, which overlap with green technologies. This extension is also useful for identifying the wide range of classification codes adopted in digital technologies, which is an input for the strategy presented in the next subsection. We draw keywords from scientific reports and academic publications focusing on the intersection between green and digital technologies (see Appendix A for an overview of extracted keywords per source). The literature generates a starting set of keywords related to digital and/or sustainable technologies. Then, we collect the abstracts and titles of patents containing a combination of both a digital and a sustainability keyword. We generated samples of this dataset to analyse the accuracy of keywords in identifying digital sustainability technologies. This accuracy is secured by the specialists' evaluation of the patents identified though each keyword by reading each patent's title and abstract. We excluded keywords retrieving poor matches. We apply a threshold of 10% as the maximum acceptable limit of false positives. Keywords that do not retrieve any digital sustainability technology are also excluded (see Appendix A).

Table 17: Keywords selected for identifying digital and sustainable technologies.

Digital-related keywords		**Sustainability-related keywords**	
3D print*	fog computing	air polluti*	GHG reduc*
adaptive robotic*	industry 4.0	biodiversity	green energy
augmented reality	information system*	biofuel*	greenhouse gase* reduction
autonomous	intelligence	carbon footprint	recycling
big data	intelligent	circular economy	reduc* carbon emission
blockchain	internet	clean energy	reduc* pollution
business analytic*	machine learning	clean fuel*	reduc* resource consumption
chip technolog*	mobile comput*	climate change	renewable energy
cloud comput*	natural language process*	climate disaster*	resource efficien*
cyber	quantum computing	CO2 emission*	resource-efficien*
data analytic*	smart	CO2 level*	green cit*
data transmission	software	eco-friend*	smart farming
data-based	traffic optimi*	electric vehicle*	solid waste
digital	virtual	energy consumption reduct*	sustainability
digitization		environmental protect*	waste management
distributed comput*		environmental-friendly	water efficiency
eBanking		environmentally-friendly	water leakage
eCommerce		food waste*	water management
eHealth		pollution control	water scarcity
eLearning		pollution detect*	water treatment

Source: Authors.

We selected a total of 34 and 40 keywords to identify digital and sustainable technologies respectively (see Table 17). We apply the '*' signal for selected keywords to capture variations of the terms. In this way, the keyword 'mobile comput*', for example, would capture all patents that have the keywords 'mobile computers', 'mobile computing', or 'mobile computation', besides any other possible similar variations occurring after the '*'. We conduct the keyword-based searches as not case-sensitive and consider both titles and abstracts of patents (i.e., registers are retrieved if they match the keyword criteria in the title or abstract).

Much of the extant research linking digital and sustainable technologies focuses on Artificial Intelligence (see for example WEF, 2018; OECD, 2019; Goh, 2021). It stresses a potentially large role for AI as a key technology in the twin transition. Thus, we adopt an additional set of keywords to improve the identification of AI-related technologies by using a combination of strategies proposed in the literature (Leusin et al., 2020; WIPO, 2019; and Cockburn et al., 2018). We implement a search using these AI keywords in an additional search step (Module 4) to identify AI technologies that are also classified under the sustainability code Y02. More information about all 128 keywords considered in these strategies and the results from their implementation is presented in Appendix B.

1.3 Classification-based strategy

As a third strategy, we aim at identifying codes that are typically used to classify digital technologies. We first identify digital technologies using the 'Digital-related keywords' (see previous Table 17). By searching for these keywords on patents' titles and abstracts, we build a dataset of 1,358,688 'digital' priority7 patents. We use this digital dataset to identify the codes typically applied to classify digital technologies. However, the usage of such codes is unbalanced, meaning that some are applied ubiquitously to classify a wide variety of patents, while others are more specific. Thus, some codes may have large absolute numbers because they are ubiquitously applied, rather than because they are particularly linked to a specific technology. Therefore, just considering the absolute number of codes related to a technology is not sufficient to identify the most important ones.

To overcome this problem, we consider the relative use of each classification code through the application of the Revealed Technological Advantage (RTA) index. Balassa (1965) proposed the Revealed Comparative Advantage (RCA) index to measure the comparative advantage of countries in the exporting of manufactured goods. It has been widely applied in the economics literature to understand how selected entities (e.g., countries, companies, industries, etc.) perform in comparison to a relative global average in the output of products. Soete (1987) adapted the specialisation measure to technologies and proposed the Revealed Technology Advantage (RTA) index. For example, Breschi et al. (2003) use the RTA index to show that companies typically patent in codes technologically close to the ones in which they hold some kind of technological advantage.

We use the RTA index to measure how selected entities perform in relation to a global average. To identify codes typically used in digital technologies, we suggest comparing the relative use of each possible classification code from our digital dataset to the average usage of the very same code in a larger economy. In this case, the RTA index can be calculated as follows:

$$\text{RTA}_{classification\ code\ c,\text{entity e}} = \frac{\dfrac{\text{Number of patents classified with code c linked entity e}}{\text{Total number of codes linked to entity } e}}{\dfrac{\text{Total number of patents classified with code } c \text{ in a larger economy E}}{\text{Total number of codes linked to the larger economy E}}}$$

If an entity has an RTA equal to or higher than one in a given classification code, this entity has a specialisation in this code, whereas values below this threshold indicate the absence of specialisation. In the current approach, the entity considered is a generic technology (i.e., digital technologies). However, we need to determine 'the larger economy' as well. One possibility is to consider all patents (i.e., digital ones plus all remaining patents ever registered) as being the larger economy. Thereby, the RTA highlights codes that are relatively 'overused' in digital technologies in comparison to their average use. Another possibility is to define the larger economy based on geography8. This approach has the advantage of insights from the literature on economic complexity, which highlight that geographies have particular technological profiles (Hidalgo, 2021). This allows using these differences to identify particularities of the entity of interest. For the implementation of such approach, one has also to define which level of geographical aggregation is most appropriate (e.g., geographies defined at the level of cities, regions, states, countries, economic blocks, etc.). We suggest that the best geographical delimitation for this approach is the country level, since the total number of patents in digital technologies is comparable to the number of patents for selected countries9.

We test both options, i.e. entity based on technology as well as entity based on geography, by identifying which one highlights classification codes that are better related to digital technologies. We do so by reading the description of every code identified using the two each options. The analysis documents that both possibilities yield very similar results (see Section 2.3 in Part II). Due to simplicity of implementation, we chose the technology-based option. Accordingly, our RTA-based strategy compares the relative use of each classification code found in the digital dataset (i.e., a number between 0 and 1) to the relative use of these very same code in all existing patents (i.e., also a number between 0 and 1). If a particular code is relatively rarely used in all patents, but it is relatively highly used in digital patents, this code is identified as being 'typical' of digital technologies. The proposed identification of codes typically

7 For comparison, if non-priorities are included in this 'digital' dataset, the total number of patents would be 2,398,495 (rather than 1,358,688). The 3,931,281 sustainability patents identified through code Y02, and 63,120 digital patents identified through code Y04, in turn, contain 1,555,434 and 30,486 priorities, respectively. Only priorities are considered when calculating the RTA index.

8 Ideally, it would be the aggregate averages of several other technologies intrinsically different from each other and that hold a number of patents comparable to the number associated with digital technologies. But circumventing one single technology from others is already a challenge, which makes such identification of several circumvented 'comparable' alternatives impractical.

9 The distribution in the number of patents each country holds is very skewed. The Top 20 countries hold an average of 1,571,307 priority patents each, whereas the following 178 countries hold on average 3,878 priority patents each (using fractional count).

associated with digital technologies reduces false positives in the final dataset, compared to an identification based on the absolute number of codes.

In this search module, we use the IPC classification, since it has a broader coverage regarding adoption across distinct patent offices compared to the CPC classification. IPC codes can be translated to CPC at a later stage, in case there is a need to do so. The IPC scheme offers five hierarchical levels (see Table 18). We first test at which level the 'particularities' of digital technologies are better highlighted (i.e., at which level the codes identified using the RTA approach are more or less 'typical' of digital technologies).

Table 18: Example of hierarchical levels adopted in the IPC scheme for the code G06N 3/02.

Hierarchical level	Code	Description
1. Section	G	Physics
2. Class	G06	Computing; Calculating; Counting
3. Subclass	G06N	Computer systems based on specific computational models
4. Group	G06N 3	Computer systems based on biological models;
5. Subgroup	G06N 3/02	Using neural network models

We do so by analysing the results of applying the RTA at each of these distinct hierarchies. Taking the broader 'Section' and the more specific 'Subgroup' level as examples, this means analysing which codes are highlighted as being typical of digital technologies: i) when taking the 8 Sections as input in the RTA equation, and ii) when the 74,503 Subgroups10 are considered as input in the RTA equation. These two calculations result in Sections G and H being identified as typical of digital technologies, and 8,431 Subgroups being identified respectively. After applying this classification for the other three possible hierarchies and analysing the results, we chose 'Groups' and 'Subclasses' as the most appropriate hierarchy levels for this particular strategy.

The RTA index is known for overestimating codes seldomly used (Soete, 1987). This seldom use makes the denominator of the RTA index too small, making the RTA results very sensitive to small numbers in the numerator. To account for this, we apply an additional threshold to each considered hierarchy (i.e., Groups and Subgroups), so that we pick just the highest RTA values. The codes related to these highest values are evaluated individually through the reading of their descriptions. Without applying the thresholds, 11% (892 out of 8,205 possible) and 14% (92 out of 653 possible) of the IPC Groups and Subclasses, respectively, show a positive RTA. From these, we select the 4311 and 10 of the highest scoring Groups and Subclasses, respectively, that were identified as related to digital technologies through the reading of their descriptions (see Section 2.5 of Part II for a list of the identified codes, including the ones excluded after the reading of their descriptions).

[10] This value considers the number of Subgroups available in the 2019 IPC scheme considered in the PATSTAT 2019 version. For statistics of IPC, see https://www.wipo.int/classifications/ipc/en/ITsupport/Version20190101/transformations/stats.html.

[11] The implementation of the RTA-technological option returns 42 codes, whereas the geographical options return 43 codes. All codes from the technological options are found in the geographical option, which just highlight additionally the code "G06E 1". To get a larger output, we opt for now to include this additional code to the 42 found in the technological option.

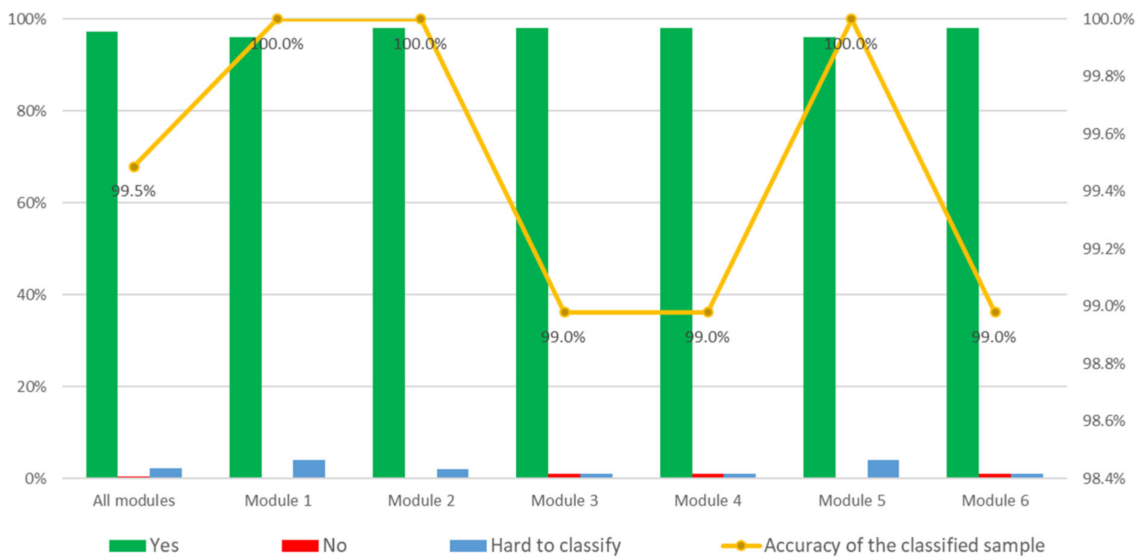
2. VALIDITY OF SEARCH RESULTS

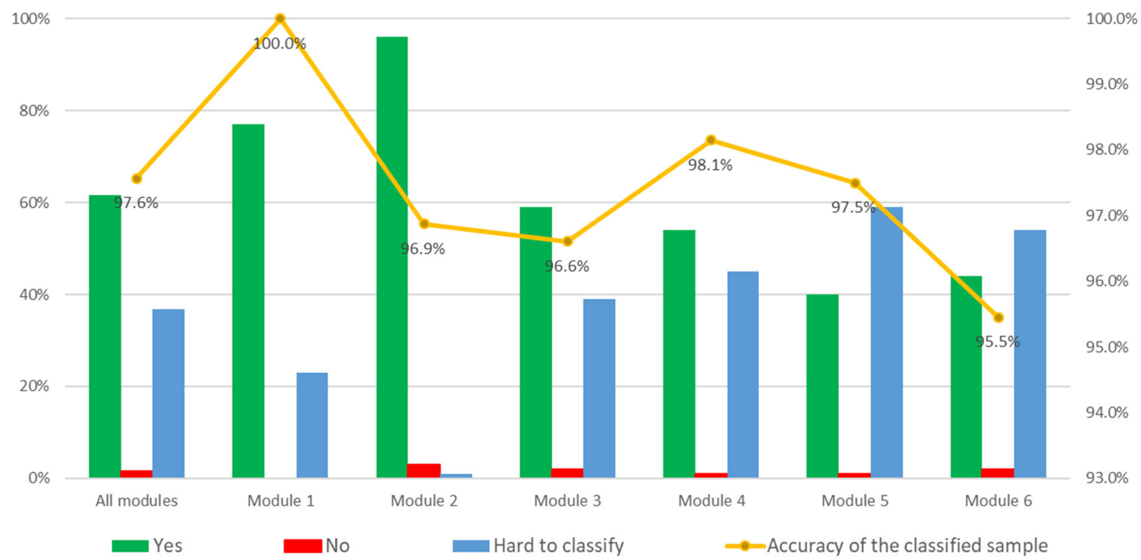
2.1 Overview of modules' performance

The most important criterion to check the quality of the identification approach is through reading the specific information of the patents identified. To this end, we created six random samples of 100 patents each to evaluate the accuracy of each of the proposed search modules. We read abstract and title of each of the 600 randomly selected patents and classified them according to being related to a technology that is digital AND sustainable, or not (see Figure 13). We report patents classified as related to digital technologies (see 13a) and patents classified as sustainable (see 13b). The reported accuracy refers to the number of true positives in comparison to false positives.

The reported accuracy is above 95.5% for all modules considering both the digital and sustainable aspects. Whereas title and abstract are often enough to classify a patent as being related to a digital technology (12a), it is not straightforward to do the same for sustainable technologies (12b). This is because the sustainable aspect is much more dependent on the interpretation of the potential of the technology (conversely, the digital aspect is a technical characteristic of a technology being related to computation, for example). The only exception to this is Module 2, since the abstracts and titles of these patents necessarily have a sustainability-related keyword. With the exception of Module 2, all remaining considered modules use the Y02 sustainable code as a proxy for identifying sustainable technologies. This means that the quality of the sustainable aspect is directly linked to the quality of the Y02 classification, which has a reported error rate of less than 7% (Angelucci et al. 2018). Considering this and the average accuracy of all modules of 99.5% in capturing digital technologies (for the considered samples), we estimate that between 92.5% and 99.5% of the patents that compose the final dataset are related to digital sustainability technologies.

Figure 13: Number of patents classified in regards to being related to a digital (13a) and to a sustainable (13b) technology.

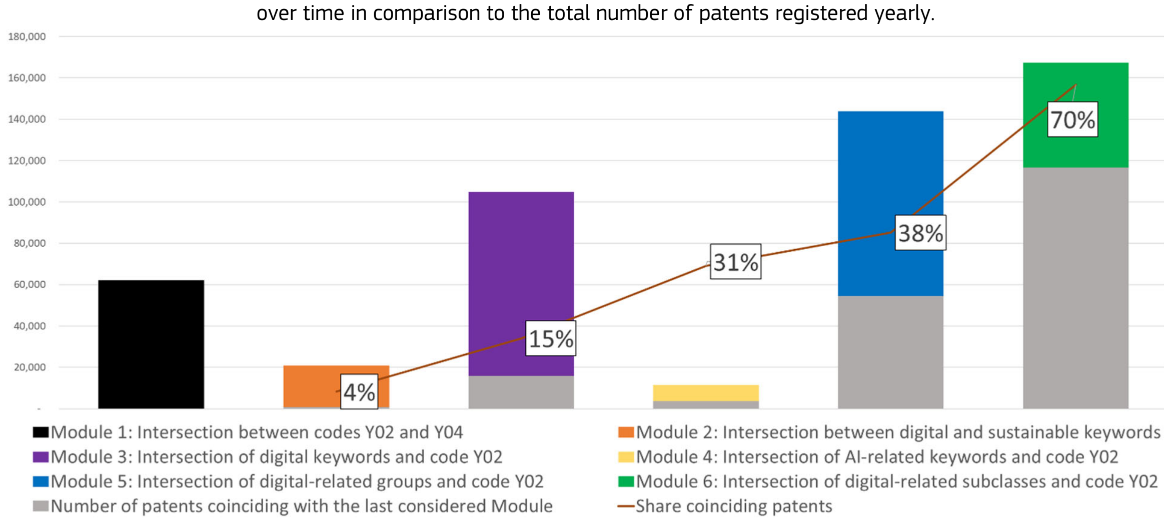


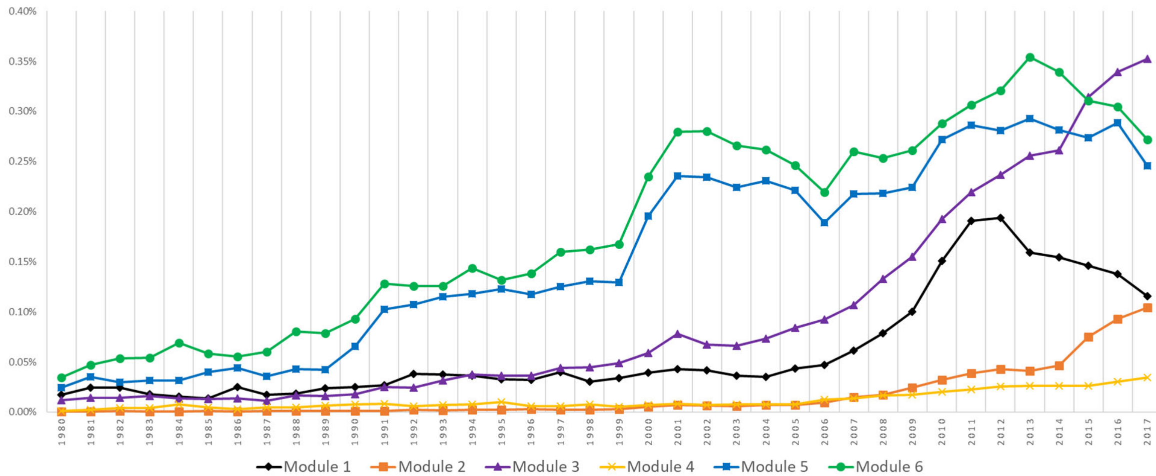


Note: Reported accuracy = 1 – (number of patents wrongly selected by the sample/number of patents correctly selected).
Source: Authors' calculations.

In addition, we evaluate the quality of the adopted identification strategy for digital sustainability inventions through an analysis of selected indicators that may reveal inconsistencies related to the considered modules. First, we consider the number of patents identified by each module (see Figure 14). Thereby, we reveal the share of patents overlapping from one module to the next (14a), and the evolution of the number of patents identified by each module since the early 1980s in comparison to the total number of patents registered yearly (14b).

Figure 14: Number of patents identified and overlapping in each Module (13a), and share of patents registered in each Module over time in comparison to the total number of patents registered yearly.





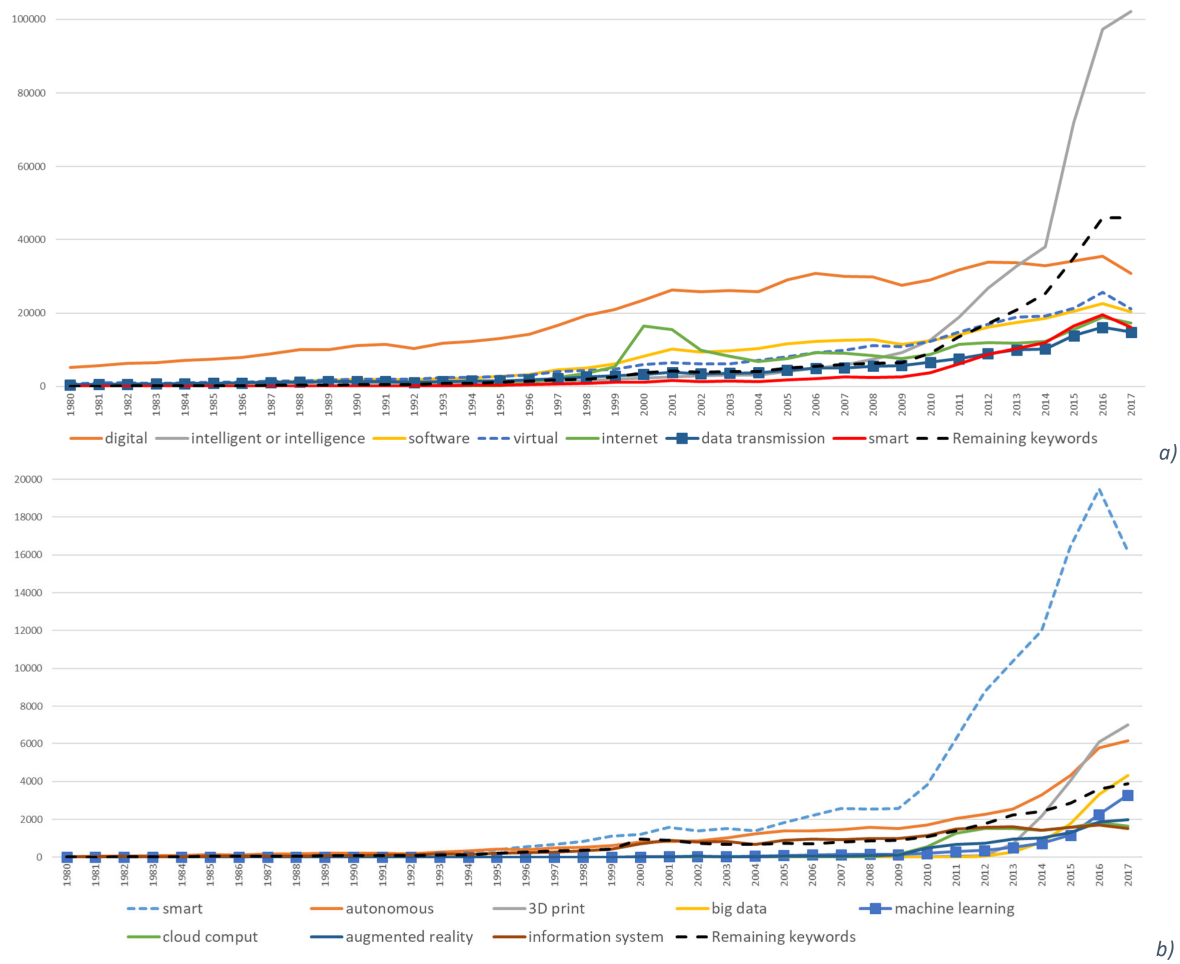
Note: due to the larger number of patents identified under the Y02 code, this code in particular is plotted on the secondary Y axis (in green).
Source: PATSTAT 2019a.

We detect no major inconsistency or deviation based on the considered indicators, although we find some particularities. The keywords-based approach of Module 2 has very little overlap with the Y-based approach adopted in Module 1 (see Figure 14a). Conversely, there is a high overlap for the number of patents identified using the subclass-based approach (Module 6). The overlap is also relatively high for Modules 4 and 6. For Module 4, in particular, this is partially due to the fact that some of its keywords also appear in the list of digital keywords considered in Module 2 (like 'intelligence' in Module 2, with 'Artificial Intelligence' in Module 4; the latter is entirely included in the former through the single term 'intelligence').

The yearly outputs of the modules based exclusively on classification codes (i.e., Modules 1, 5, and 6) have been decreasing since 2013, whereas the outputs of modules based on keywords (Modules 2, 3, and 4) have been increasing particularly after 2014 (see Figure 14b). This might indicate a 'trending' effect of some keywords over time. To check for trending effects of the selected keywords, we analyse the use of each of the 34 digital keywords12 suggested previously in Table 17 (see Figure 15). We highlight the adoption of the 7 most used digital-keywords in comparison to the remaining 27 keywords Figure 15a) and further disaggregate remaining keywords, separating the 8 most used among these from the remaining 19 digital-keywords (see Figure 15b).

[12] The focus on digital-keywords over sustainable keywords is justified by their usage in more modules (namely in Modules 2 and 3, whereas sustainable keywords are used just in Module 2). Module 4 also is based on some keywords that are directly related to digital technologies (e.g., machine learning and, as highlighted above, a reference to the keyword "intelligence").

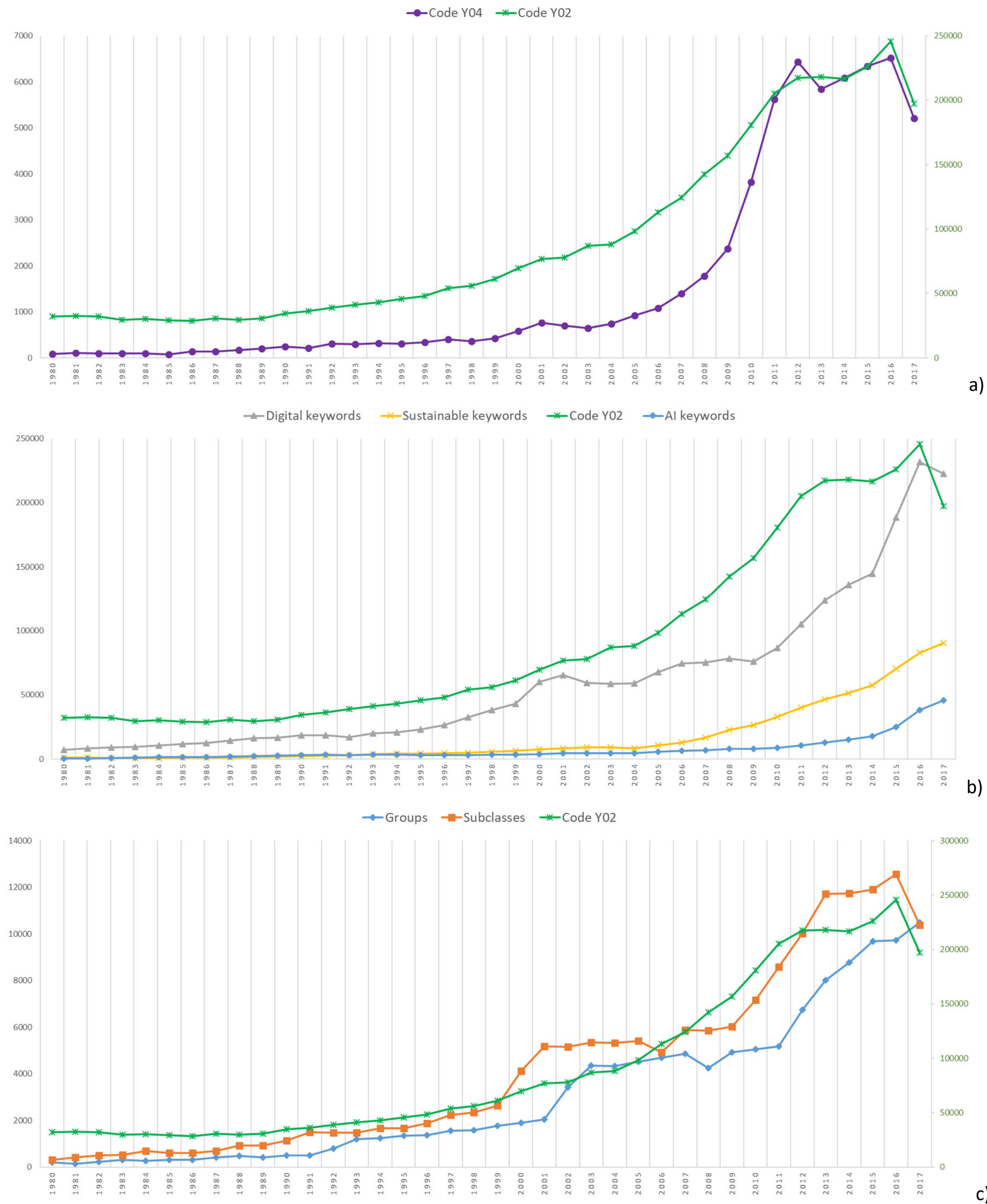
Figure 15: Number of patents captured by the 7 most used keywords (15a), and from the following 8th to the 15th most used keywords (15b).



Source*: PATSTAT 2019a*.

Apart from the keywords 'digital', 'internet' and 'software', the remaining top 4 performing digital keywords have grown in usage mostly after 2009 (see Figure 15a). Between these, the largest share of digital patents is captured by the keywords 'intelligent' or 'intelligence'. Similarly, from the remaining 8 highly used keywords, the highest share of patents is captured by a keyword 'smart' (see Figure 15b). We observe also emerging technologies in the second set of most used keywords, including keywords such as '3D print', 'big data', 'machine learning', 'cloud computing', and 'augmented reality' growing after 2010. Thus, the increase in the number of patents captured by the use of keywords highlighted before is associated with trending of specific keywords.

Figure 16: Number of Y02 patents in comparison to the number of patents identified under the Y04 code (16a), patents identified using the digital and sustainability keywords (16b), and patents identified through the selected IPC groups and subclasses (16c).

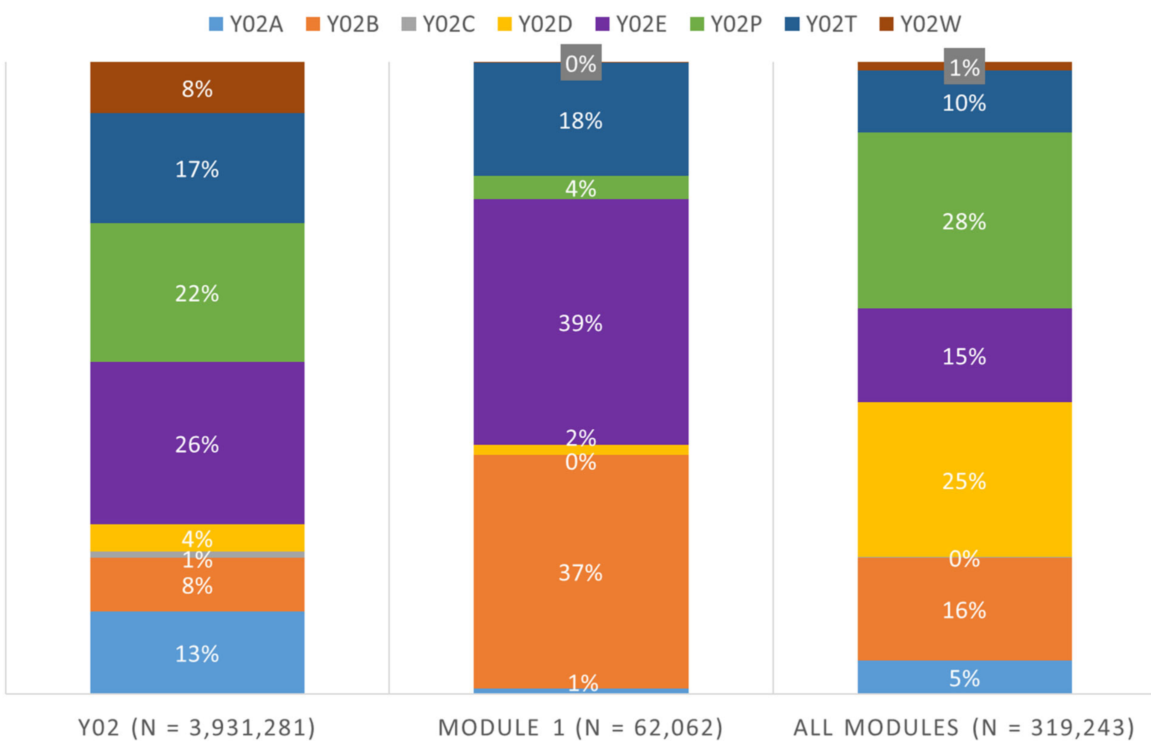


Note: We plot the number of patents identified under the Y02 code on the secondary Y axis in 16a and 16c.
Source: PATSTAT 2019a.

Next, we compare the number of patents identified by the Y02 'green' code to i) the number of patents identified in the Y04 digital code (16a); ii) the number of patents identified using digital and sustainability keywords (16b); and iii) the number of patents identified using the selected IPC groups and subclasses (17c). We observe in the performance of each strategy in comparison to the Y02 code, that the subclasses-based strategy is the one that more closely follows the trend seen for the sustainable technologies identified with the Y02 code (see Figure 16c). Code Y04, in particular, shows a faster development after 2008 (16a), whereas for the digital-related keywords the increase is steady, with two 'dents' in the trend observed at the beginning of the 2000s and after 2008 (16b). In comparison to the other searches, the patents identified using sustainability keywords show a particularly lower output, despite the high number of keywords adopted for this identification.

Next we consider how the extension of the search through the addition of distinct modules affected the composition of the complete dataset in comparison to the search based exclusively on Y codes (Module 1). We look into how the codes available in the Y02 classification (as provided previously in Table 16) change when all proposed modules are considered in comparison to the composition of Module 1. We use the composition of all patents classified under the Y02 code as a reference for this comparison (see Figure 17).

Figure 17: Distribution of patents across subclasses in a dataset with all Y02 patents, Module 1 and all Modules.



Source: PATSTAT 2019a.

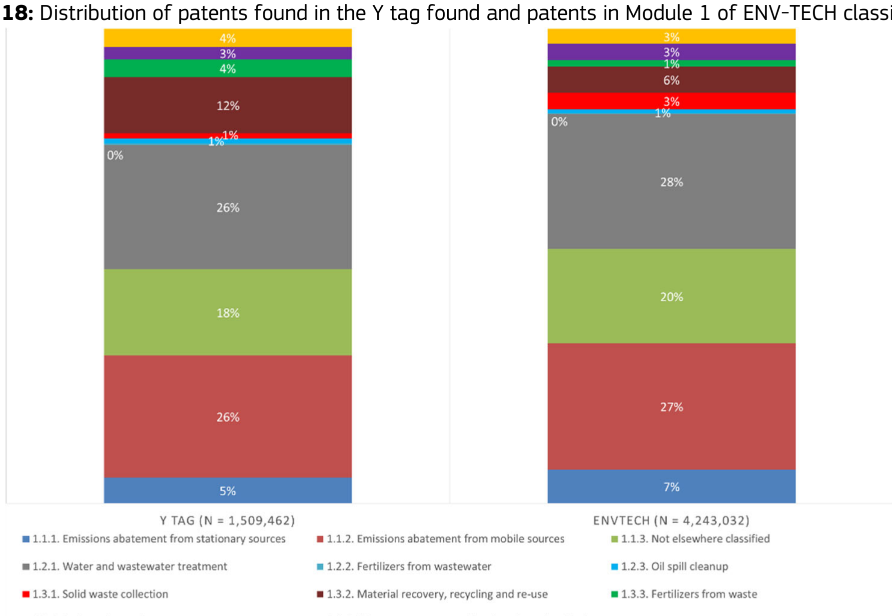
We observe that the exclusive consideration for the Y04 code as a proxy for identifying digital technologies, implemented in Module 1, results in a high share of patents related to the sustainable Y02B (which refers to patents linked to 'Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications'), whereas the Y02P is not strongly represented (related to 'Climate change mitigation technologies in the production or processing of goods'). The over and underrepresentation of the Y02B and Y02P subclasses, respectively, can be explained by the fact that the Y04 code refers exclusively to 'Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grids'. Such systems are an integral part of buildings and house appliances (Y02B) and are not typically found in the production of processing of goods (Y02P). When widening the scope of what is considered a digital technology through the additional modules, the underrepresented Y02P increases its relative share and getting closer to the distribution of the Y02 code, whereas the overrepresentation of the Y02E declines. We also observe a reduction in the relative shares for the codes Y02W ('Climate change mitigation technologies related to wastewater treatment or waste management') and Y02A ('Technologies for adaptation to climate change') in the final dataset in comparison to the Y02 distribution. These relative reductions are compensated by a large increase of the relative share of the code Y02D ('Climate change mitigation technologies in information and communication technologies [ICT], i.e. information and communication technologies aiming at the reduction of their own energy use'). This seems plausible given that ICTs (Y02D) are an intrinsic part of digital technologies, whereas technologies related to energy generation/transmission/distribution (Y02E), waste treatment (Y02W), and climate change (Y02A) are expected to be considerably more representative for sustainable technologies than for digital ones.

2.2 ENV-TECH classification vs. the 'Y' CPC classification

The Environment-Related Technologies (ENV-TECH) classification as proposed by Hascic and Migotto (2015) builds upon past work from the OECD Working Party on Integrating Environmental and Economic Policies and the Working Party on Climate, Investment and Development, as well as over the 'Y' CPC classification. It has 9 modules: Environmental Management (Module 1); Water-Related Adaptation Technologies (Module 2); Biodiversity Protection and Ecosystem Health (Module 3); Climate Change Mitigation Technologies Related to Energy Generation, Transmission or Distribution (Module 4); Capture, Storage, Sequestration or Disposal of Greenhouse Gases (Module 5); Climate Change Mitigation Technologies Related to Transportation (Module 6); Climate Change Mitigation Technologies Related to Buildings (Module 7); Climate Change Mitigation Technologies Related to Wastewater Treatment or Waste Management (Module 8); and Climate Change Mitigation Technologies in the Production or Processing of Goods (Module 9). Particularly, Module 3 is proposed but not yet implemented in the ENV-TECH classification, whereas Modules 4 to 9 are exclusively based on the 'Y' classification. Module 4 is entirely based on the Y subclassification code Y02E, Module 5 is based on the code Y02C, Module 6 on code Y02T, Module 7 on code Y02B, Module 8 on code Y02W, and Module 9 on code Y02P. Non-Y-related Modules, i.e., Modules 1 and 2, apply a variety of IPC classification codes to identify relevant patents. The implementation of the strategies proposed in these two IPC-based modules generates 2,488,871 registers (PASTAT 2019a). Modules 4 to 9, in turn, provide additional 3,295,471 patents. In total, the ENV-TECH strategy identifies 5,784,342 patents. For comparison, the alternative 'Y' classification scheme identifies 3,994,401 patents in total, which is roughly 31% fewer patents than the ENV-TECH.

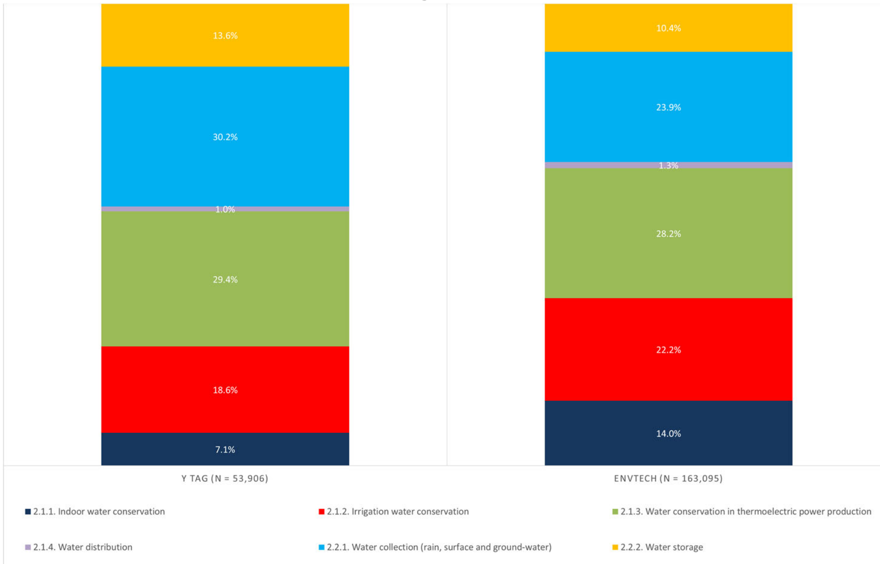
We find that 82.5% of all patents identified in the 'Y' classification scheme are also in the ENV-TECH classification. There is also an overlap when the patents exclusively found in the ENV-TECH strategy are considered: 29.5% of all patents identified in Modules 1 or 2 from the ENV-TECH classification are also found in the Y tag. The distribution of these overlapping patents is remarkably similar (see Figures 18 and 19).

Figure 18: Distribution of patents found in the Y tag found and patents in Module 1 of ENV-TECH classification.



Source*: PATSTAT 2019a.*

Figure 19: Distribution of patents found in the Y tag and patents in Module 2 of the ENV-TECH classification.

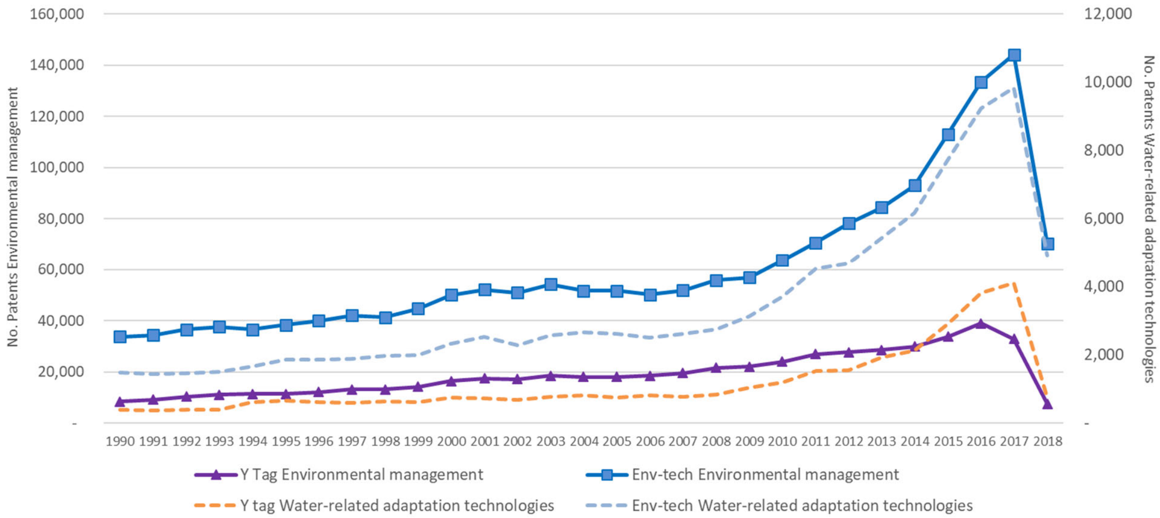


Source: PATSTAT 2019a.

We find no significant difference in the trends over time by considering the number of registers over the years for Modules 1 and 2 in comparison to the patents captured with the Y tag (see Figure 20). This similarity in compositions and time trends indicates that the Y tag does not favour any particular code from the ones considered in Modules 1 and 2. However, there are differences when one considers the Y codes that are neglected in the ENV-TECH classification, which does not consider the subclasses 'Technologies for adaptation to climate change' (Y02A) and 'Climate change mitigation technologies in ICTs' (Y02D), and 'Information or communication technologies having an impact on other technology areas' (Y04)13. To understand how this affects the results, the IPC codes from the neglected Y-tagged patents are compared to the patents identified exclusively in Modules 1 and 2 from the ENV-TECH classification. In total, the Y codes Y02A, Y02D, and Y04 capture 827,325 patents, whereas Modules 1 and 2 capture 2,488,871 patents.

[13] The subclass Y02A captures 574,257 patents, subclass Y02D captures 187,653 patents, and the code Y04,captures 72,630 patents.

Figure 20: Registers over the years for all patents identified in Module 1 (Environmental management) and 2 (Water-related adaptation technologies) in comparison to the patents from these modules that are also found under the Y tag.



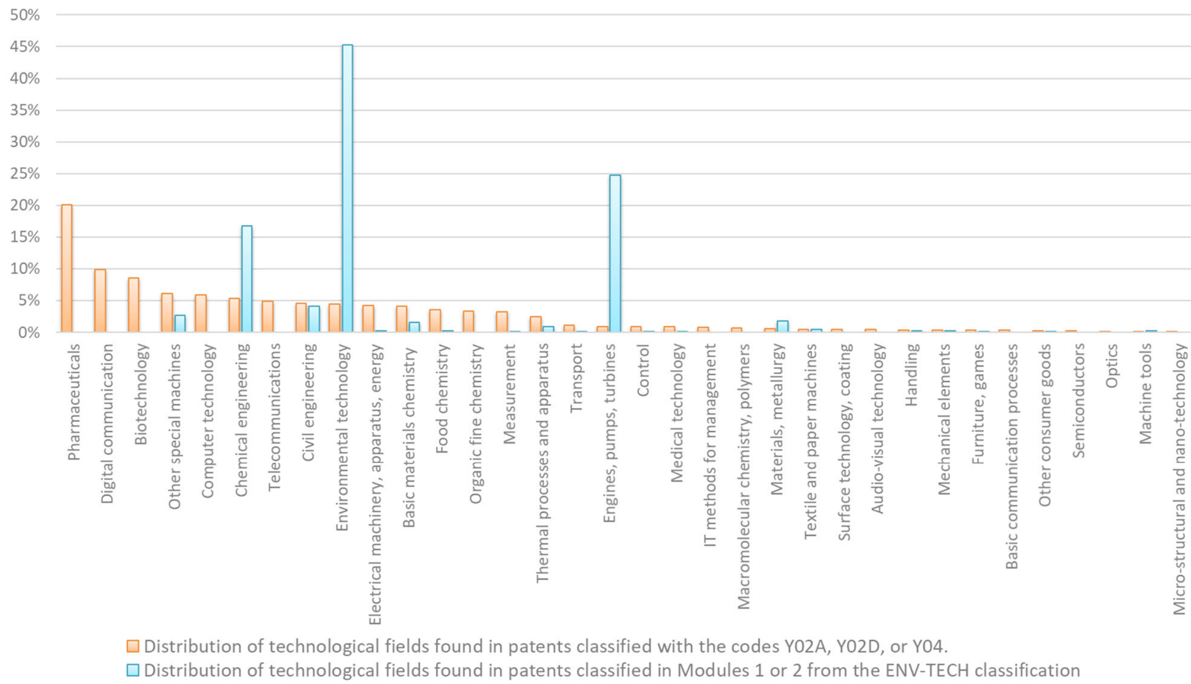
Source*: PATSTAT 2019.*

The first insight from the comparison is, how broader the Y scheme is regarding the number of IPC subclasses found: there are 623 unique IPC subclasses linked to Y dataset, in comparison to 64 unique IPC subclasses found in the selected ENV-TECH dataset. This broad variety of IPC codes makes it difficult a direct comparison of both datasets. To overcome this, the IPC subclasses are converted to their equivalent technological fields in line with Schmoch (2008). There are 35 technological fields in this scheme, from which 34 are found in at least one of the two considered datasets (see Figure 21).

The large majority (85.4%) of technological fields related to patents found in Modules 1 or 2 refer to the fields of 'Environmental technology', 'Chemical engineering', or 'Engines, pumps, turbines'. Particularly, the high representativeness of the technological field 'Environmental technology' can be linked to the fact that 55.6% of the subclasses linked to this field in Schmoch (2008) are also considered in Modules 1 or 2[14]. The distribution is broader for the patents found in the Y codes Y02A, Y02D, and Y04: 13 technological fields (all fields from 'Pharmaceuticals' until 'Organic fine chemistry', from the left to the right in Figure 21) must be considered to get to a similar representativeness. Together these 13 technological fields account for 85.0% of the registers found. Modules 1 and 2 overlook digital technologies, since we find no patents in the ENV-TECH dataset linked to the technological fields of 'Digital communication', 'Computer technology', and 'Telecommunications', whereas in the Y dataset these three fields are linked to 20.7% of the registers.

[14] Namely the IPC subclasses B01D, B65F, F01N, F23G, F23J are considered in these said modules, whereas the subclasses A62D, G01T, E01F, A62C are not.

Figure 21: Composition of the selected Y and ENV-TECH datasets considering the technological fields linked to their patents.



Source: PATSTAT 2019.

Hence, we can conclude that both the ENV-TECH classification and the Y classification yield a similar output, while the patent recall of the latter is roughly 31% lower. The majority of the patents classified in the Y scheme (82.5%) are also found in the ENV-TECH dataset due to the fact that the latter is largely based on Y codes, since 6 out of 8 Modules of the ENV-TECH classification are based exclusively on Y codes. The differences exists when one compares the Y codes not part of the ENV-TECH classification to the classification-based Modules of the latter approach (i.e., Modules 1 and 2). In the Y selected dataset we find a higher number of different IPC subclasses and better coverage of digital technologies linked to the technological fields of 'Digital communication', 'Computer technology', and 'Telecommunications'. Therefore, the 'Y' classification scheme seems preferable when one is particularly interested in identifying technologies that combine both a sustainability and a digital component with the trade-off of having a lower output. In addition, the Y classification offers has a quality filter, applied through the specialists' opinion, which is likely to reduce its output compared to the ENV-TECH dataset, whilst increasing the accuracy through less false positives.

2.3 Testing RTA-Indices in the classification-based search modules

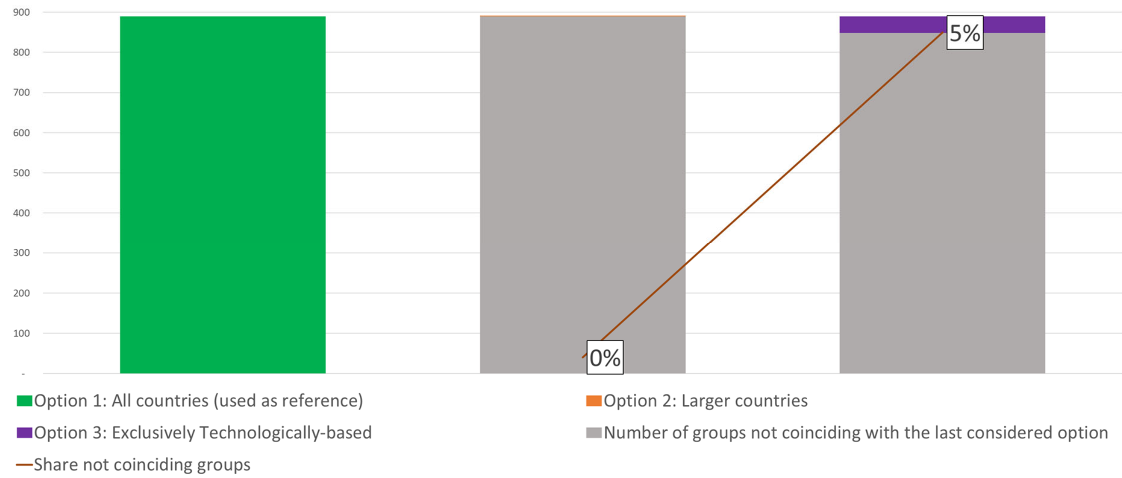
We test three possible options of 'larger economies' in the application of RTA-Indices as part of our classification-based search module: two are based on geography and one technologies.

- Option 1: We consider all countries which results in 299 patent offices.
- Option 2: We exclude smaller countries from option 1 and use the 39 largest patent offices.
- Option 3: We uses the average use of each patent classification code.

For the geography-based options, we follow De Rassenfosse et al. (2019) and assign patents to countries by using patents inventors' addresses. We test each option at the Group and Subclass IPC level. The results for both hierarchies show that the three options overlap to a great extent. 892 Groups are classified as specialised in the first two options, whereas 914 Groups are classified in the same way in Option 3. For simplicity, we exclude the 22 lowest performing Groups from Option 3, so that we can compare the same number of codes (892) for the three options. For subclasses, the three options classify as specialised the same number of 92 codes. When comparing these codes, one finds that

890 out of the 892 Groups for Options 1 and 2 coincide, whereas 848 groups from Option 3 coincide with Option 1 or 2. For subclasses, all 92 codes identified as specialised are the same across the three considered options. Figures 22 and 23 highlight this overlap for IPC Groups and Subclasses, respectively.

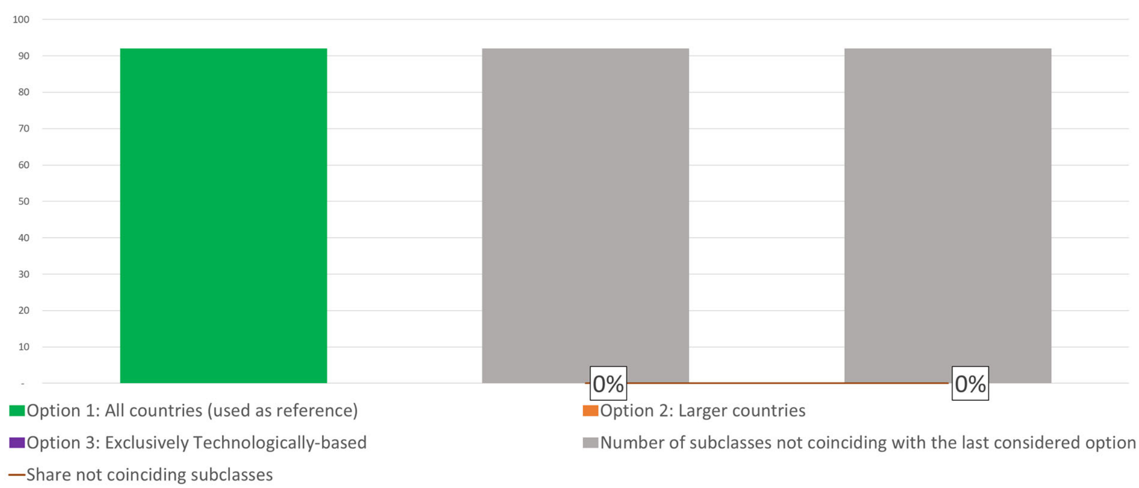
Figure 22: Number of coinciding Groups for the three considered options.



Source: PATSTAT 2019a.

We further classify, by reading the codes description, the highest scoring 110 groups and 80 subclasses of each option. We classify these codes into being related to digital technologies, or not. This analysis shows that 51.8% of the highest scoring groups for Options 1 and 2 are confirmed as being related to digital technologies, whereas this share is 51.2% for Option 3. For subclasses, 50.7% of the highest scoring codes are identified as related to digital technologies for the three options.[15]

Figure 23: Number of coinciding subclasses for the three considered options.



Source: PATSTAT 2019a.

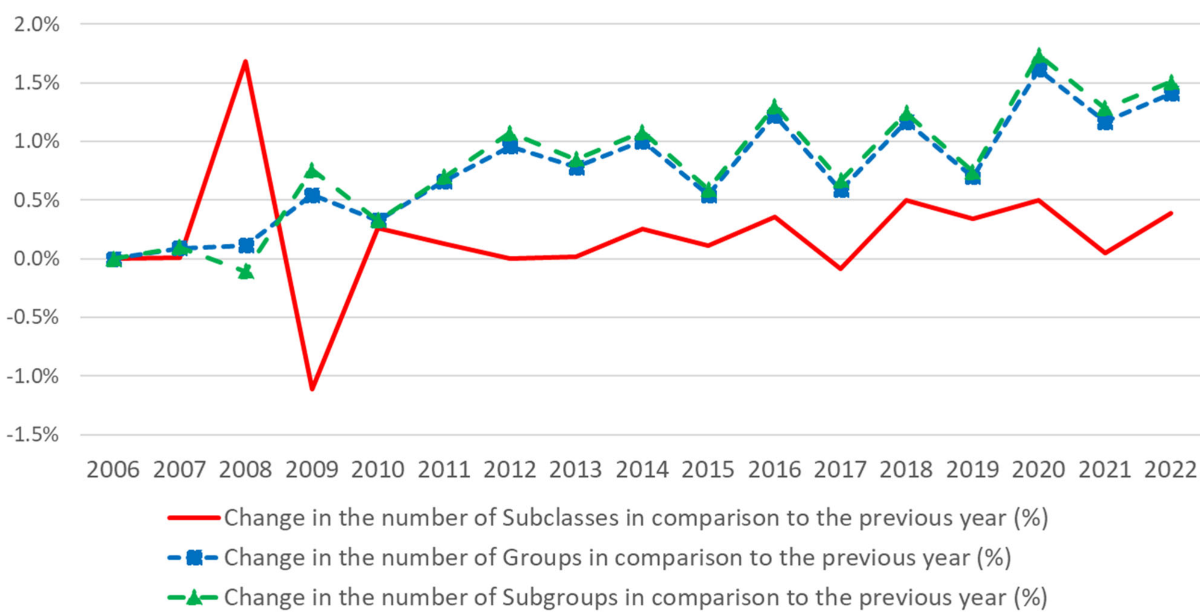
[15] More details of this analysis are included in the files shared with the JRC team: 'Groups_Description_analysis' and 'Subclasses_Description_analysis'

2.4 Changes in IPC codes over time

We also check the possible effects that updates of the IPC classification scheme may have on the codes chosen in the search strategy. The IPC classification has been updated regularly since 2006 (WIPO, 2022)16. These updates include the inclusion and exclusion of selected codes, and modifications in the description of codes. This can potentially affect Modules 5 and 6 of the proposed identification strategy, since some of the identified Subclasses and Groups may refer to codes that were modified. To analyse the possible extension of this issues, we consider two distinct indicators related to changes in IPC codes: i) changes over time in the number of existing IPC codes at distinct hierarchy levels, and ii) the kind of modifications implemented in each year17.

Figure 24 highlights the first indicator by showing the changes in the number of Subclasses, Groups, and Subgroups in comparison to the number of codes existing in the previous year. Although Subgroups are not considered in the proposed strategy, they are presented here as a reference due to their higher specificity. It seems that the difference is marginal compared to the Group hierarchy. Except for 2008 for Subgroups, and 2009 and 2017 for Subclasses, the number of codes considered in the IPC classification for these three hierarchies has been steadily increasing over time. For the whole 16-year period (2007-2022), there was an increase of 3.0%, 12.5%, and 11.6% in the total number of Subclasses, Groups, and Subgroups, respectively. This suggests that IPC codes are relatively stable over time. It also shows that the exclusion of codes is less frequent than the creation of new ones. This, in turn, indicates that the codes selected in Modules 5 and 6 are less prone to be excluded over time, which supports the replication of the adopted search.

Figure 24: Relative changes over the years in the number of codes applied in the IPC classification for selected hierarchies (Subclasses, Groups, and Subgroups).



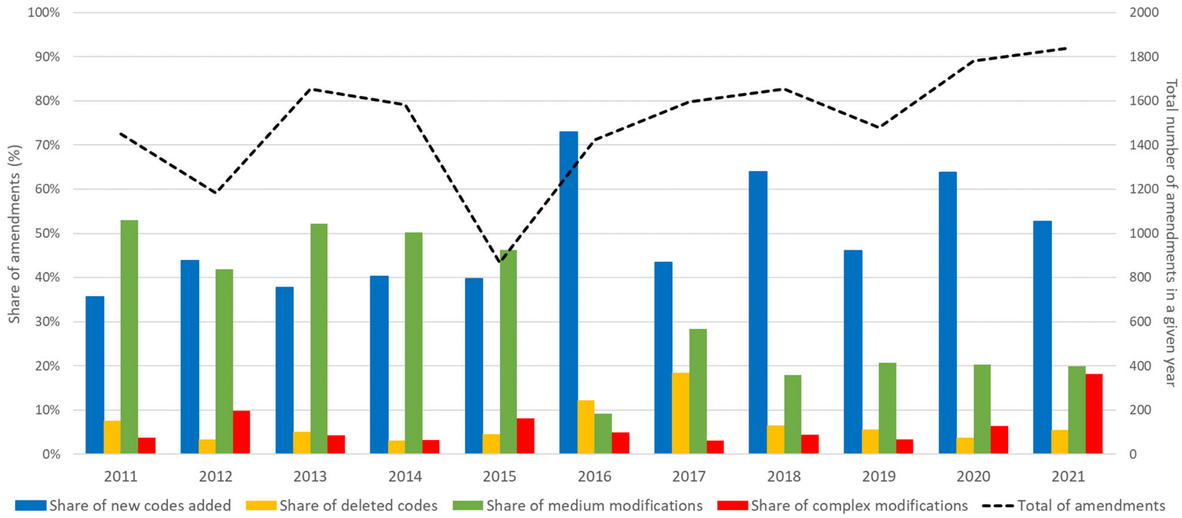
Sources: WIPO (2021, 2022) and authors' calculations.

Next, we investigate the kind of amendments that have been applied to the IPC codes over time. Figure 25 disaggregates these numbers in comparison to the total number of amendments per year for the English language. The standard currently used for disaggregating amendments was adopted in 2011 and finishes in 2021.

16 This happened on an annual basis with some exceptions, e.g., the years 2007 and 2008 had two updates. Before 2006, changes were more seldom. There are 7 updates in total presented in WIPO (2022) for the period before 2006.

17 See https://www.wipo.int/ipc/itos4ipc/ITSupport_and_download_area/20210101/stats/IPC_statistics-20210101_V2.0.zip for the first indicator, and https://www.wipo.int/classifications/ipc/en/ITsupport/old_versions.html for the second.

Figure 25: Kinds of amendments applied to the IPC classification over time.



Sources: WIPO (2021, 2022) and authors' calculations.

The majority of amendments indeed refers to the addition of new codes, followed by 'medium modifications', which refer to changes in the description of the codes (see Figure 25). The average annual changes are 49.1%, 32.7%, 6.9%, and 6.3% for the share of new codes added, share of medium modifications, share of deleted codes, and share of complex modifications, respectively. Assuming that medium modifications do not impact sufficiently individual codes to the extent of being used to classify distinct technologies18, the changes that may impact the codes selected in Modules 5 and 6 refer to deletions and complex modifications. As these changes are just a fraction of the total changes, we consider that there has been not a significant number of codes, used in Modules 5 and 6, modified over time. Given that 43 groups and 10 subclasses are proposed in Modules 5 and 6, respectively, there is a chance of 0.016% (0.00036%*43) for groups and 0.009% (0.00091%*10) for subclasses that at least one of the respective proposed codes was (or will be, since we are considering 2018 data) deleted or modified in the period from 2006 to 2022. These figures increase to 0.022% and 0.013% for groups and subclasses, respectively, when one extrapolates the assumptions to include the 7 IPC updates made before 2006.

[18] Unfortunately, there is no precise information given in WIPO (2022) regarding what "medium" and "complex" modifications mean.

2.5 Groups and subclasses identified using the RTA-based strategy[19]

Table 19: IPC Groups identified as typically adopted to classify patents related to digital technologies.

Group	RCA	Description	Seemingly related to Digital technologies?
G04C 99	23.74	Subject matter not provided for in other groups of this subclass	Hard to classify
G02F 7	14.59	Optical analogue/digital converters	Yes
H03M 1	14.46	Analogue/digital conversion; Digital/analogue conversion (conversion of analogue values to or from differential modulation H03M0003000000)	Yes
H03M 3	13.70	Conversion of analogue values to or from differential modulation	No
B33Y 40	13.57	Auxiliary operations or equipment, e.g. for material handling	No
B29C 64	13.02	Additive manufacturing, i.e. manufacturing of three-dimensional [3D] objects by additive deposition, additive agglomeration or additive layering, e.g. by 3D printing, stereolithography or selective laser sintering	Hard to classify
B33Y 70	13.02	Materials specially adapted for additive manufacturing	Hard to classify
B33Y 99	12.95	Subject matter not provided for in other groups of this subclass	Hard to classify
H03H 17	12.71	Networks using digital techniques	Yes
B33Y 30	11.96	Apparatus for additive manufacturing; Details thereof or accessories therefor	Hard to classify
G06J 3	11.87	Systems for conjoint operation of complete digital and complete analogue computers	Yes
H03M 99	11.87	Subject matter not provided for in other groups of this subclass	Hard to classify
H04H 3	11.87		Invalid code
G06T 19	11.16	Manipulating 3D models or images for computer graphics	Yes
G06J 1	11.16	Hybrid computing arrangements	Yes
B33Y 50	11.08	Data acquisition or data processing for additive manufacturing	Yes
G05D 27	8.73	Simultaneous control of variables covered by two or more of main groups G05D0001000000-G05D0025000000	Hard to classify
H03B 28	8.67	Generation of oscillations by methods not covered by groups H03B0005000000-H03B0027000000, including modification of the waveform to produce sinusoidal oscillations	No
B33Y 80	8.36	Products made by additive manufacturing	Hard to classify
B33Y 10	8.14	Processes of additive manufacturing	Hard to classify
G05B 15	8.01	Systems controlled by a computer	Yes
G06N 99	7.86	Subject matter not provided for in other groups of this subclass	Hard to classify

[19] The highest scoring Groups and Subclasses identified as related to digital technologies are presented in Tables 18 and 19, respectively. These codes are filtered through their descriptions, remaining 43 Groups and 10 Subclasses (i.e., all codes that are not classified with a 'yes' in the column 'seemingly related to digital technologies' are excluded). The remaining codes are used in the fifth (Group-based) and sixth (Subclass-based) modules proposed.

G07F 15	7.67	Coin-freed apparatus with meter-controlled dispensing of liquid, gas, or electricity	No
H04M 7	7.42	Arrangements for interconnection between switching centres	No
G16H 80	7.32	ICT specially adapted for facilitating communication between medical practitioners or patients, e.g. for collaborative diagnosis, therapy or health monitoring	Yes
H04N 101	7.22	Still video cameras	No
G08C 17	7.01	Arrangements for transmitting signals characterised by the use of a wireless electrical link	Yes
H04L 29	6.97	Arrangements, apparatus, circuits or systems, not covered by a single one of groups H04L0001000000-H04L0027000000	Hard to classify
G16H 40	6.87	ICT specially adapted for the management or administration of healthcare resources or facilities; ICT specially adapted for the management or operation of medical equipment or devices	Yes
H04N 21	6.80	Selective content distribution, e.g. interactive television or video on demand [VOD]	No
H02J 13	6.78	Circuit arrangements for providing remote indication of network conditions, e.g. an instantaneous record of the open or closed condition of each circuit breaker in the network; Circuit arrangements for providing remote control of switching means in a power distribution network, e.g. switching in and out of current consumers by using a pulse code signal carried by the network	Yes
H03L 9	6.78	Automatic control not provided for in other groups of this subclass	Yes
H04H 5	6.78		Invalid code
H04M 99	6.78	Subject matter not provided for in other groups of this subclass	Hard to classify
G08B 19	6.66	Alarms responsive to two or more different undesired or abnormal conditions, e.g. burglary and fire, abnormal temperature and abnormal rate of flow	Yes
H04B 14	6.64	Transmission systems not characterised by the medium used for transmission	Hard to classify
H04H 60	6.57	Arrangements for broadcast applications with a direct linkage to broadcast information or to broadcast space-time; Broadcast-related systems	Yes
H04H 20	6.51	Arrangements for broadcast or for distribution combined with broadcast	Hard to classify
G07C 9	6.51	Individual registration on entry or exit	Hard to classify
G05B 17	6.44	Systems involving the use of models or simulators of said systems	Yes
H04H 7	6.33		Invalid code
H04J 7	6.21	Multiplex systems in which the amplitudes or durations of the signals in individual channels are characteristic of those channels	No
G09C 5	6.19	Ciphering or deciphering apparatus or methods not provided for in other groups of this subclass, e.g. involving the concealment or deformation of graphic data such as designs, written or printed messages	Yes
H04J 9	6.15	Multiplex systems in which each channel is represented by a different type of modulation of the carrier	No
H04L 27	6.10	Modulated-carrier systems	No
F24F 130	6.06	Control inputs relating to environmental factors not covered by group F24F 110/00	Yes
G16H 20	6.04	ICT specially adapted for therapies or health-improving plans, e.g. for handling prescriptions, for steering therapy or for monitoring patient compliance	Yes
G16H 15	6.04	ICT specially adapted for medical reports, e.g. generation or transmission thereof	Yes
G06F 9	6.03	Arrangements for program control, e.g. control units	Yes
G01R 22	6.01	Arrangements for measuring time integral of electric power or current, e.g. electricity meters	No
G06Q 99	5.99	Subject matter not provided for in other groups of this subclass	Hard to classify

G06F 8	5.96	Arrangements for software engineering	Yes
B63B 69	5.93	Equipment for shipping not otherwise provided for	No
G16H 50	5.89	ICT specially adapted for medical diagnosis, medical simulation or medical data mining; ICT specially adapted for detecting, monitoring or modelling epidemics or pandemics	Yes
H04H 40	5.85	Arrangements specially adapted for receiving broadcast information	No
H04W 80	5.84	Wireless network protocols or protocol adaptations to wireless operation	No
H03M 5	5.80	Conversion of the form of the representation of individual digits	No
G06T 17	5.79	3DÂ modelling for computer graphics	Yes
G05B 19	5.77	Programme-control systems	Hard to classify
H03H 21	5.75	Adaptive networks	Yes
G04G 21	5.73	Input or output devices integrated in time-pieces	Hard to classify
G08C 23	5.69	Non-electric signal transmission systems, e.g. optical systems	No
H04L 12	5.68	Data switching networks	Yes
H04Q 11	5.64	Selecting arrangements for multiplex systems	No
G09B 5	5.62	Electrically-operated educational appliances	No
H04L 25	5.61	Baseband systems	No
G06F 21	5.60	Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	Yes
F24F 120	5.50	Control inputs relating to users or occupants	Yes
H04L 7	5.46	Arrangements for synchronising receiver with transmitter	No
G01R 13	5.45	Arrangements for displaying electric variables or waveforms	Yes
G06Q 30	5.45	Commerce, e.g. shopping or e-commerce	Yes
H04H 9	5.41		Invalid code
H04L 9	5.38	Arrangements for secret or secure communication	Yes
G06T 13	5.37	Animation	Yes
G06Q 20	5.37	Payment architectures, schemes or protocols	Yes
G04G 9	5.32	Visual time or date indication means	No
A43D 33	5.28	Machines for assembling lifts for heels	No
G16H 30	5.25	ICT specially adapted for the handling or processing of medical images	Yes
H04J 3	5.21	Time-division multiplex systems	No
G04G 17	5.15	Structural details; Housings	No
H04N 7	5.14	Television systems	No

G11B 20	5.00	Signal processing not specific to the method of recording or reproducing; Circuits therefor	No
G01R 35	4.99	Testing or calibrating of apparatus covered by the other groups of this subclass	No
H04K 1	4.90	Secret communication	Hard to classify
H04L 1	4.89	Arrangements for detecting or preventing errors in the information received	Yes
H04W 40	4.86	Communication routing or communication path finding	No
G06Q 50	4.79	Systems or methods specially adapted for specific business sectors, e.g. utilities or tourism	Yes
G09B 9	4.77	Simulators for teaching or training purposes	Yes
H99Z 99	4.75	Subject matter not otherwise provided for in this section	Hard to classify
G09F 27	4.72	Combined visual and audible advertising or displaying, e.g. for public address	No
H04J 4	4.70	Combined time-division and frequency-division multiplex systems	No
G01D 21	4.65	Measuring or testing not otherwise provided for	Hard to classify
H04W 12	4.64	Security arrangements; Authentication; Protecting privacy or anonymity	Yes
G06Q 10	4.64	Administration; Management	Yes
G01C 11	4.61	Photogrammetry or videogrammetry, e.g. stereogrammetry; Photographic surveying	No
G06T 15	4.56	3D [Three Dimensional] image rendering	Yes
G05D 1	4.55	Control of position, course, altitude, or attitude of land, water, air, or space vehicles, e.g. automatic pilot	Yes
H04W 84	4.54	Network topologies	No
H04H 1	4.54		Invalid code
G06F 15	4.52	Digital computers in general; Data processing equipment in general	Yes
G07C 11	4.52	Arrangements, systems or apparatus for checking, e.g. the occurrence of a condition, not provided for elsewhere	No
H03L 7	4.50	Automatic control of frequency or phase; Synchronisation	No
H04L 5	4.49	Arrangements affording multiple use of the transmission path	No
H04B 11	4.48	Transmission systems employing ultrasonic, sonic or infrasonic waves	No
H04L 23	4.46	Apparatus or local circuits for telegraphic systems other than those covered by groups H04L0015000000-H04L0021000000	No
G01D 4	4.46	Tariff metering apparatus	No
G06E 1	4.44	Devices for processing exclusively digital data	Yes

Source: Authors.

Table 20: IPC Subclasses identified as typically adopted to classify patents related to digital technologies.

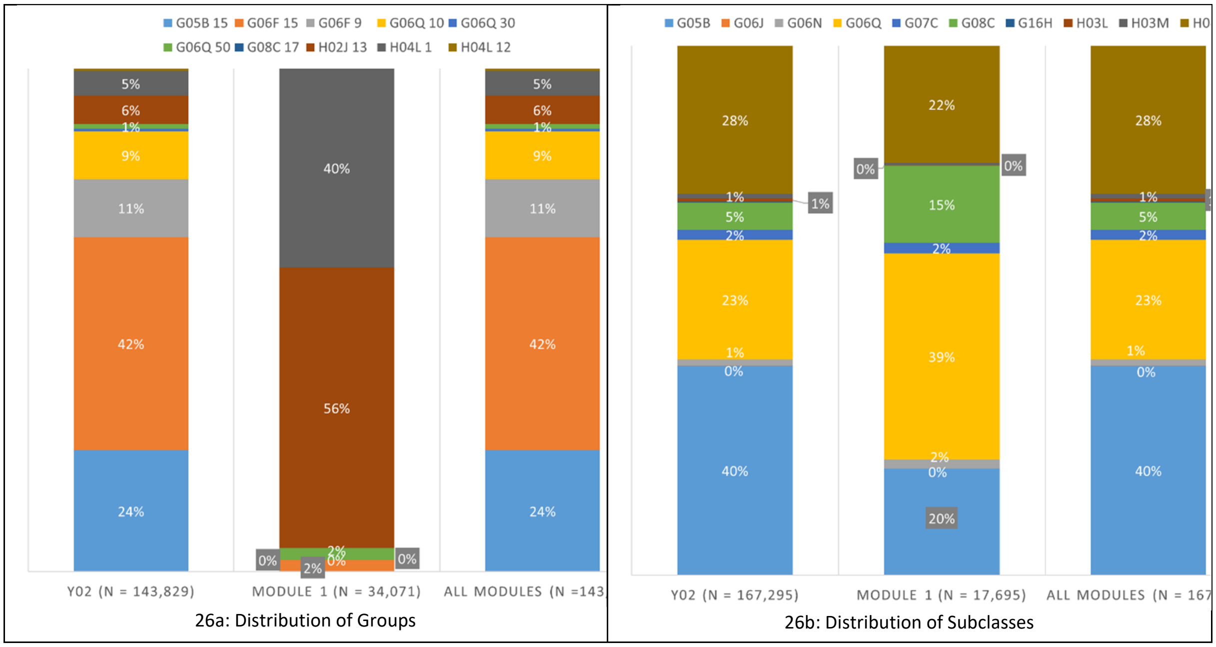
Subclass	RCA	Description	Seemingly related to Digital technologies?
B33Y	11.54	Additive manufacturing, i.e. Manufacturing of three-dimensional [3d] objects by additive deposition, additive agglomeration or additive layering, e.g. By 3d printing, stereolithography or selective laser sintering	No
G06J	10.89	Hybrid computing arrangements	Yes
H03M	8.13	Coding, decoding or code conversion, in general	Yes
H04H	6.20	Broadcast communication	No
H04L	5.93	Transmission of digital information, e.g. Telegraphic communication	Yes
G08C	5.48	Transmission systems for measured values, control or similar signals	Yes
G16H	5.36	Healthcare informatics, i.e. Information and communication technology [ict] specially adapted for the handling or processing of medical or healthcare data	Yes
G05B	4.99	Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements	Yes
G07C	4.95	Time or attendance registers; registering or indicating the working of machines; generating random numbers; voting or lottery apparatus; arrangements, systems or apparatus for checking not provided for elsewhere	Yes
G06Q	4.85	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for	Yes
H99Z	4.75	Subject matter not otherwise provided for in this section	Hard to classify
H04K	4.61	Secret communication; jamming of communication	Hard to classify
G06N	4.46	Computer systems based on specific computational models	Yes
H03L	4.33	Automatic control, starting, synchronisation, or stabilisation of generators of electronic oscillations or pulses	Yes

Source: Authors.

2.6 Groups and subclasses in Module 1 in comparison to Y02 classification

We also check how the composition of the digital sustainability patents collected using the proposed strategy in comparison to alternative strategies adopting exclusively Module 1 or Y02 codes. Figures 26a and 26b highlight the representativeness of these three strategies in the 10 most used Groups and Subclasses, respectively, considered in Modules 5 and 6.

Figure 26: Representativeness of the 10 most used Groups/Subclasses in Modules 5 and 6.



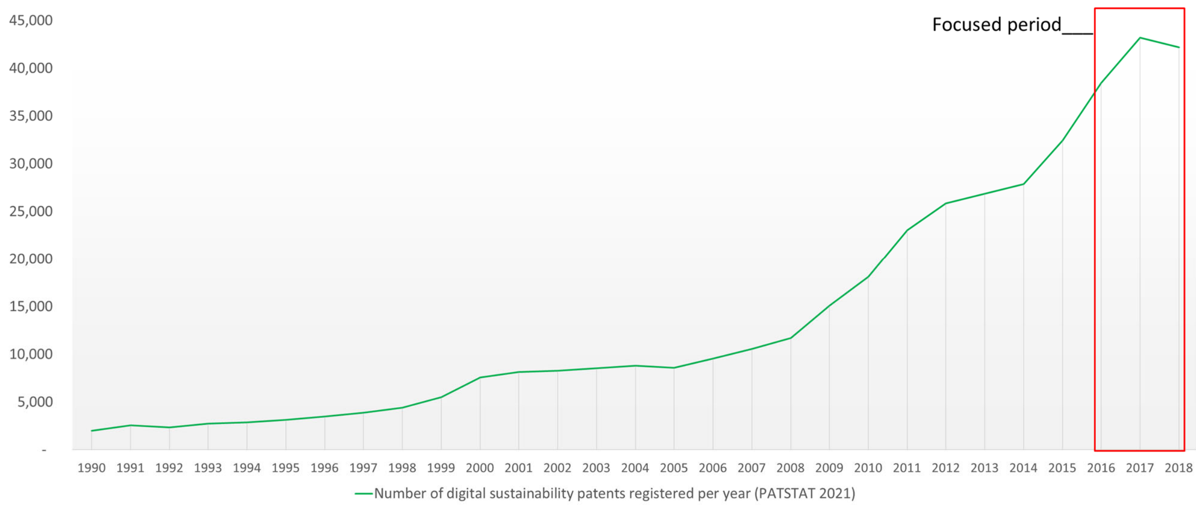
Notes: Since Modules 5 and 6 use all of these codes as criteria for retrieving all patents tagged with the Y02 code, the composition of the final results using all modules is identical to the composition of patents found exclusively under the Y02 code.
Source: PATSTAT 2019.

The top 10 Groups represent 88.7% of patents in Module 5, whereas this share is 100% for subclasses considered in Module 6 i.e., the 10 subclasses considered comprehend all codes adopted in this module. Module 1 has an over representation of groups linked to electricity (section H, e.g., group H02J 13, related to 'Circuit arrangements for providing remote indication of network conditions, etc.', and group H04L 1 related to 'Arrangements for detecting or preventing errors in the information received'. Groups linked to computation (e.g., group G05B 15 and G06F 15) are underrepresented. For subclasses, there is over representation of codes linked to the distribution of electricity (e.g., subclass G06Q and G08C). Subclasses from Module 1 are underrepresented in subclass G05B. This finding confirm the exclusive focus of the Y04 code (Module 1) on technologies related to electric power. Whereas subclasses H02J 13 and H04L 1 are potentially linked to the aspect of operating electric generation and distribution, subclasses G06Q and G08C seem to be linked more broadly to forecasting and supervision. All modules jointly, in turn, generate a dataset with even distribution of groups and subclasses. For groups, in particular, a higher focus on computing technologies is evident (e.g., group G06F 15 linked to 'Digital computers in general; Data processing equipment in general'), and the already mentioned group G05B 15 related to 'Systems controlled by a computer', and group G06F 15 related to 'Digital computers in general').

2 MATCHING WITH DATA FOR TOP R&D INVESTORS

In the next section, we investigate the contribution towards the development of digital sustainability inventions by the world's top corporate R&D investors. To this end, we apply the proposed identification strategy to PATSTAT 2021b. This update not only increases the number of digital sustainability patents identified to 494,255, it also allows improved coverage of more recent years, which a selected for the firm level analysis (see Figure 27).

Figure 27: Number of digital sustainability patents registered between 1990 and 2018 updated to PATSTAT 2021.



*Source***:** PATSTAT 2021.

We link patents to R&D Scoreboard companies using the JRC-OECD COR&DIP© v.3 dataset (Hernández et al. 2020; Amoroso et al. 2021). To this end, a list of inventors related to the R&D scoreboard companies is first identified in the JRC-OECD COR&DIP© v.3 dataset (as of 2019) and then used as a reference to establish patents' ownership. However, this list of inventors is static over time, meaning that it does not consider the possibility of inventors entering or leaving R&D scoreboard companies over the years. Given that, R&D Scoreboard companies have large, complex and especially changing ownership structures due to exits, M&As, etc., we need to restrict the firm level analysis to a relatively short period of 3 years (2016- 2018). We adopt the same list of identified inventors to define patent ownership over the same period of analysis. The earliest filling year of each priority is considered for the definition of the year in which a patent was created.

We identify priority patents using the IP5 strategy for the period from 2016 to 2018. This strategy considers only the patents that were registered at least in one of the 5 largest IP offices20 and in at least two distinct patent offices worldwide (i.e., single filings are excluded). Using the IP5 strategy excludes patents filled in just one patent office and generates a dataset of related inventions with a 'comparable' technological and economic value (Dernis et al., 2015).

Accordingly, instead of using only the PATSTAT's DOCDB simple family identifier (i.e., 'docdb_family_id') as done in Section 2 of Part I of this report, the dataset is limited in Section 3 (Part I of the report) to patent families that have at least one register in the IP5. Thus, single filings are excluded. This is implemented by considering only the priorities registered with a "prior_appln_id" in PATSTAT's "TLS204_APPLN_PRIOR" table. We also exclude utility models (application kind 'U'), Plant patents (application kind 'V'), and Design patents (application kind 'F').

[20] Namely European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (CNIPA) and the United States Patent and Trademark Office (USPTO)

REFERENCES

Amoroso, S., Aristodemou, L., Criscuolo, C., Dechezleprete, A., Dernis, H., Grassano, N., Moussiegt, L., Napolitano, L., Nawa, D., Squicciarini, M., 2021. World Corporate Top R&D investors: Paving the way for climate neutrality. EUR.

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

Balassa, B., 1965. Trade liberalisation and "revealed" comparative advantage. The manchester school 33, 99-123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

Balland, P.A., Boschma, R., Crespo, J., Rigby, D.L., 2019. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. Regional studies 53, 1252-1268. https://doi.org/10.1080/00343404.2018.1437900

Breschi, S., Lissoni, F., Malerba, F., 2003. Knowledge-relatedness in firm technological diversification. Research policy 32, 69-87. https://doi.org/10.1016/S0048-7333(02)00004-5

Castaldi, C., Frenken, K., Los, B., 2015. Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. Regional studies 49, 767-781. https://doi.org/10.1080/00343404.2014.940305

Cockburn, I.M., Henderson, R., Stern, S., 2018. The Impact of Artificial Intelligence on Innovation, The Economics of Artificial Intelligence: An Agenda. National Bureau of Economic Research. https://doi.org/10.3386/w24449

De Rassenfosse, G., Kozak, J., Seliger, F., 2019. Geocoding of worldwide patent data. Scientific data 6, 1-15. https://doi.org/10.1038/s41597-019-0264-6

Dernis, H., Dosso, M., Hervas, F., Millot, V., Squicciarini, M., Vezzani, A., 2015. World corporate top R&D investors: Innovation and IP bundles. Joint Research Centre (Seville site).

Foray, D., David, P.A., Hall, B., 2009. Smart specialisation–the concept. Knowledge economists policy brief 9, 100.

Foray, D., David, P.A., Hall, B.H., 2011. Smart specialisation from academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation. EPFL. MTEI-WORKING_PAPER-2011-001

George, G., Merrill, R.K., Schillebeeckx, S.J., 2021. Digital sustainability and entrepreneurship: How digital innovations are helping tackle climate change and sustainable development. Entrepreneurship Theory and Practice 45, 999-1027.

Goh, H.-H., 2021. Artificial Intelligence in Achieving Sustainable Development Goals. arXiv preprint arXiv:2107.13966.

Hascic, I., Migotto, M., 2015. Measuring environmental innovation using patent data. OECD Environment Working Papers, 0_1.

Hernández, H., Grassano, N., Tübke, A., Amoroso, S., Csefalvay, Z., Gkotsis, P., 2020. The 2019 EU Industrial R&D Investment Scoreboard. EUR 30002 EM. Publications Office of the European Union. doi 10, 04570.

Hernández, H., Grassano, N., Tuebke, A., Potters, L., Gkotsis, P., Vezzani, A., 2018. The 2018 EU industrial R&D investment scoreboard, EU R&D Scoreboard Report. doi: 10.2760/131813, JRC113807

Hidalgo, C.A., 2021. Economic complexity theory and applications. Nature Reviews Physics 3, 92-113.

Inaba, T., Squicciarini, M., 2017. ICT: A new taxonomy based on the international patent classification.

Leusin, M.E., Günther, J., Jindra, B., Moehrle, M.G., 2020. Patenting patterns in Artificial Intelligence: Identifying national and international breeding grounds. World Patent Information 62, 101988. https://doi.org/10.1016/j.wpi.2020.101988

Montresor, S., Quatraro, F., 2020. Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies. Regional studies 54, 1354-1365. https://doi.org/10.1080/00343404.2019.1648784

OECD, 2019. Measuring the Digital Transformation. doi:https://doi.org/10.1787/9789264311992-en

Russel, F., 2021. Industry Classification Benchmark (ICB). FTSE Russel

Schmoch, U., 2008. Concept of a technology classification for country comparisons. Final report to the world intellectual property organisation (wipo), WIPO.

Soete, L., 1987. The impact of technological innovation on international trade patterns: the evidence reconsidered. Research policy 16, 101-130. https://doi.org/10.1016/0048-7333(87)90026-6

Stuermer, M., 2014. Characteristics of digital sustainability, Proceedings of the 8th international conference on theory and practice of electronic governance, pp. 494-495.
USPTO, 2021. CPC Section Y, in: USPTO, E.a. (Ed.), https://www.uspto.gov/web/patents/classification/cpc/html/cpc-Y.html.

Veer, T., Jell, F., 2012. Contributing to markets for technology? A comparison of patent filing motives of individual inventors, small companies and universities. Technovation 32, 513-522. https://doi.org/10.1016/j.technovation.2012.03.002

WEF, 2018. Harnessing artificial intelligence for the earth. Fourth Industrial Revolution for the Earth Series.

WIPO, 2019. WIPO Technology Trends 2019: Artificial Intelligence, Geneva: World Intellectual Property Organization.

WIPO, 2021. IPC statistics 2021. https://www.wipo.int/ipc/itos4ipc/ITSupport_and_download_area/20210101/stats/IPC_statistics-20210101_V2.0.zip

WIPO, 2022. International Patent Classification (IPC) - IT support area - Old editions. https://www.wipo.int/classifications/ipc/en/ITsupport/old_versions.html

Wolfson, J.A., 1993. Patent Flooding in the Japanese Patent Office: Methods for Reducing Patent Flooding and Obtaining Effective Patent Protection. Geo. Wash. J. Int'l L. & Econ. 27, 531.

List of Abbreviations and definitions

CCMA	Climate change mitigation and/or adaptation
CNIPA	People's Republic of China National Intellectual Property Office
CPC	Cooperative Patent Classification
ENV-TECH	Environment-Related Technologies
EPO	European Patent Office
EU	European Union
EUIPO	European Union Intellectual Property Office
GHG	Greenhouse gas
ICB	Industry Classification Benchmark
ICT	Information and Communication Technology
IP	Intellectual Property
IP5	Five IP offices (EPO, JPO, KIPO, USPTO and CNIPA)
IPC	International Patent Classification
ISIC	International Standard Industrial Classification of All Economic Activities
JPO	Japan Patent Office
KIPO	Korean Intellectual Property Office
NACE	Statistical classification of economic activities in the European Community
OECD	Organization for Economic Co-operation and Development
PATSTAT	Patent Statistical Database
R&D	Research and Development
RTA	Revealed Technology Advantage
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organization
WPCID	Working Party on Climate, Investment and Development
WPIEEP	Working Party on Integrating Environmental and Economic Policies

List of Figures

List of Tables

Appendices

APPENDIX A: SOURCES FOR INITIAL KEYWORDS (TABLE A1) AND LIST OF EXCLUDED KEYWORDS (TABLE A2)

Table A1: List of scientific and technical publications considered for the choice of keywords, and keywords highlighted in each document.

Type	Publication title	Keywords digital	Keywords sustainability	Keywords that capture both digital and sustainability aspects
Paper	Digital Transformation and Environmental Sustainability: A Review and Research Agenda	artificial intelligence; big data; analytics; cloud computing; Internet of Things (IoT).	pollution control; waste management; sustainable production; urban sustainability; air pollution; CO2 emission; water treatment; climate change; disaster management; solid waste; E-waste; food waste; agri-waste; sustainable manufacturing; sustainable supply chain; smart cities; sustainable cities	-
Paper	Digital Sustainability and Entrepreneurship: How Digital Innovations Are Helping Tackle Climate Change and Sustainable Development	Internet of Things (IoT); smart houses; blockchain; Artificial Intelligence; Machine learning; big data.	-	-
Paper	Digitalization Technologies for Industrial Sustainability	artificial intelligence (AI); augmented reality (AR); IIoT and data analytics; machine learning; IIoT technology; smart energy system; factory-to-grid (F2G);	Energy consumption; energy efficiency; CO2 levels; foraging factory; re-manufacture of materials; Re-manufacturing; circular economy	-
Paper	Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level	artificial intelligence	sustainability; sustainable	digitainability
Paper	Digitainability: The Combined Effects of the Megatrends Digitalization and Sustainability	digitalization	sustainability	digitainability; sustainable digitalization
Paper	Artificial Intelligence in Achieving Sustainable Development Goals	artificial intelligence	sustainable	smart fishery; land usage monitoring; smart grid; smart city; disaster forecast; smart farming;
Paper	Assessing the Gap between Technology and the Environmental Sustainability of European Cities	big data analytics; Internet of Things (IoT); Artificial Intelligence (AI); ICT-enabled applications; eCommerce; eBanking; eHealth; eLearning;	-	-
Paper	Assessing the nexus of sustainability and information & communications technology	-	air quality; clean water & sanitation; biodiversity & habitat; forests; water resources; circular and sharing economies;	green use of robots and drones for environmental surveillance; monitoring; and protection; smart grids;
Paper	The Leading Digital Technology Companies and Their Approach to Sustainable Development	Internet of Things; cloud computing; artificial intelligence; big data analytics; smart sensors; adaptive robotics; machine	circular economy; water management; climate change; social dimensions; environment and ecosystems; waste management; reduced carbon footprint; carbon reduction;	-

		learning; smart factories; data; artificial intelligence; digital technology		
Paper	Digital transformation as a strategy to reach sustainability	artificial intelligence (AI); machine learning; cloud computing; digital transformation; bigdata; blockchain; cloud computing; social computing; augmented reality (AR); robotics; cloud computing; mobile computing; fog computing; big data; data science; business analytics; social computing; Internet of things (IoT); cyber-physical systems (CPSs);	-	-
Paper	The Effect of Blockchain Technology on Supply Chain Sustainability Performances	blockchain	Supply Chain Sustainability	-
Paper	Sustainability in the Digital Age Special Issue Introduction	Artificial intelligence (AI); big data	-	-
Paper	Is digitalization a problem solver or a fire accelerator? Situating digital technologies in sustainability discourses	big data; artificial intelligence; machine learning	-	-
Paper	Towards sustainability: PSS, digital technology and value co-creation	digital technology	sustainability	-
Report	The twin transition: a new digital and sustainability framework for the public sector	IoT; Data Collaboration; AI; Machine learning; Traffic Optimisation;	Climate disaster impact predictions & assessments; Carbon Footprint tracking & reporting; water leakage detection; Air Quality MMM (Monitor; Model; Manage); Water Efficiency; Pollution detection & recyclingt; waste collection optimization; food waste; lifecycle transparency; electrical vehicles charging	planetary computer; geospatial analytics; Energy distribution optimisation; sustainable decision-making
Report	Digital Sustainability Global sustainability as a driver of innovation and growth	digital	zero carbon mobility; resource-efficient infrastructure; renewable energy; smart nutrition; water scarcity; climate change; clean energy; destruction of biodiversity; minimize CO2 emissions; circular economy; sharing economy	-
Report	Impacts of the digital transformation on the environment and sustainability	robotics; Internet of Things (IoT); blockchain; Artificial Intelligence (AI); Augmented Reality; Blockchain; Drones; Big Data; Artificial Intelligence (AI); Automation and Robotization; Internet of Things (IoT); Autonomous driving	circular economy; clean economy; restoring ecosystems; restoring biodiversity; environmentally-friendly food system; smart farming; toxic-free environment; monitoring of biodiversity and ecosystem services; sustainable consumption; preventing food losses; monitoring and reporting of greenhouse gases (GHG) emissions; environmental offences; environmental data; reduce resource depletion; ozone depletion; terrestrial ecotoxicity; effective implementation and enforcement of environmental standards; increase resource efficiency; reduce resource consumption; reduce e-waste; automated and connected multimodal mobility;	Sustainable software; ecosystem services; biodiversity monitoration
Report	Measuring the Digital Transformation - A roadmap for the future	cloud computing; Internet of Things (IoT); artificial intelligence (AI); big data analytics; 3D printing;	-	-

Report	OECD Digital Economy - Outlook 2020	Artificial Intelligence; blockchain; quantum computing; big data analytics; semantic web; natural language processing	-	-
Report	World Economic Forum (2018): Harnessing Artificial Intelligence for the Earth	Quantum and distributed computing; Artificial intelligence;	healthy oceans; clean air; sustainable land-use; biodiversity conservation; water efficiency; clean fuels;	smart agriculture

Source: Authors.

Table A2: List of keywords excluded after analysis.

Digital keywords		**Sustainability keywords**		**Digital and sustainable related keywords**	
Keyword	Comment	Keyword	Comment	Keyword	Comment
wireless	excluded after analysis for being too generic	air quality	changed after analysis due to too many false positives	digitainability	The whole set of digital-sustainability keywords was dropped due to the low output generated (less than 20 patents are retrieved from the whole set of keywords; such a low output doesn't justify making the list of considered keywords longer through the use of these additional keywords).
		clean air	changed after analysis due to too many false positives	digital circular economy	
		environmental disaster management	changed after analysis from "disaster management" to "environmental disaster management" because 'disaster' is too generic and may not mean environmental disaster	green use of robots	
		ecosystem	changed after analysis due to too many false positives	planetary computer	
		environmental application	excluded after analysis	environmental disaster forecast	
		environmental data	excluded after analysis	-	
		environmental monitoring	excluded after analysis	-	
		environmental surveillance	excluded after analysis because 'surveillance' is too generic	-	
		habitat	changed after analysis due to too many false positives	-	
		smart cit	changed after analysis from "smart cit" to "green cit" which is more directly linked to sustainability	-	
		sustainable	changed after analysis due to too many false positives	-	
		water resource	changed after analysis due to too many false positives	-	

Source: Authors.

APPENDIX B: KEYWORDS FOR IDENTIFYING AI PATENTS

The specific keywords for the identification of AI related invention proposed in Leusin et al. (2020), WIPO (2019), and Cockburn et al. (2018) are presented in Tables B1, B2, and B3, respectively.

Table B1: List of keywords proposed in Leusin et al. (2020) for identifying AI technologies.

Keywords for identifying AI technologies as proposed in Leusin et al. (2020)		
artificial intelligence	*description logistic*	*reinforcement learn*
machine learn	*classification tree*	*multitask learn*
probabilistic reason	*regression tree*	*decision tree learn*
fuzzy logic	*logical learn*	*support vector network*
logic programming	*relational learn*	*deep structured learn*
ontology engineer	*probabilistic graphical model*	*hierarchical learn*
pervised learn	*rule learn*	*graphical model*
reinforced learn	*instance-based learn*	*structured probabilistic model*
task learn	*latent represent*	*rule induction*
neural network	*bio-inspired approach*	*memory-based learn*
deep learn	*machine intelligen*	*bio-inspired comput*
expert system	*probability logic*	*biologically inspired comput*
support vector machin	*probabilistic logic*	

Table B2: List of keywords proposed in WIPO (2019) for identifying AI technologies (Methodology report, pg. 23).

Keywords for identifying AI technologies as proposed in WIPO (2019)		
Artific intelligen%	Supervised Learning	Learning Model%
Computation intelligen%	Supervised Training%	Support Vector Machine%
Neural Network%	Supervised-Learning%	Random Forest%
Bayes network%	Swarm Intelligen%	Decision Tree%
Bayesian-Network%	Swarm-Intelligen	Gradient Tree Boosting%
Chatbot%	Unsupervised Learning	Xgboost%
Data Mining%	Unsupervised Training%	Adaboost
Decision Model%	Unsupervised-Learning%	Rankboost
Deep Learning%	Semi-Supervised Learning%	Logistic Regression
Deep-Learning%	Semi-Supervised Training	Stochastic Gradient Descent
Genetic Algorithm	Semi-Supervised-Learning	Multilayer Perceptron%
Inductive Logic Programm%	Connectionis%	Latent Semantic Analysis
Machine Learning%	Expert System%	Latent Dirichlet Allocation
Machine-Learning%	Fuzzy Logic%	Multi-Agent System%
Natural Language Generation%	Transfer-Learning	Hidden Markov Model
Natural Language Processing%	Transfer Learning	
Reinforcement Learning	Learning Algorithm%	

Table B3: List of keywords proposed in Cockburn et al. (2018) for identifying AI technologies (see Appendix A, pg. 38).

Keywords for identifying AI technologies as proposed in Cockburn et al. (2018)		
natural language processing	symbolic error analysis	computer vision
image grammars	machine learning	robot
pattern recognition	neural networks	robots
image matching	reinforcement learning	robot systems
symbolic reasoning	logic theorist	robotics
symbolic error analysis	bayesian belief networks	robotic
pattern analysis	unsupervised learning	collaborative systems
symbol processing	deep learning	humanoid robotics
physical symbol system	knowledge representation and reasoning	sensor network
natural languages	crowdsourcing and human computation	sensor networks
pattern analysis	neuromorphic computing	sensor data fusion
image alignment	decision making	systems and control theory
optimal search	machine intelligence	layered control systems
symbolic reasoning	neural network	

The three keyword-based strategies combined return a total of 295,002 unique AI related patents. The strategy proposed in Leusin et al. (2020) returns a total of 65,660 AI patents, the WIPO strategy returns 77,716 AI patents, and the strategy by Cockburn et al. (2018) returns a total of 249,819 AI patents. In contrast to Leusin et al. (2020), the other two publication do not report specific information about the accuracy of their proposed keywords for identifying AI patents. Leusin et al. (2020), report an accuracy of 95.7%, which is reportedly higher than two alternatives identification methods used by the authors for comparison (see Leusin et al. 2020, Fig. 1, pg. 5) and passes the minimum threshold (i.e., 90% accuracy) applied in the identification strategy for digital sustainability inventions. For this reason, we calculate the additions made by each strategy in comparison to the one presented in Leusin et al. (2020). Thereby, WIPO (2019) adds 20,859 AI patents to the dataset identified in Leusin et al. (2020), and Cockburn et al. (2018) adds 208,483 AI patents to the dataset identified in Leusin et al. (2020) and WIPO (2019) combined.

APPENDIX C: EXEMPLARY INVENTIONS FOR THE FOUR MAIN CLUSTERS

Table C1: Exemplary patents linked to Cluster 1.

Appln_id	Abstract	Title	CPC subcl.
477075506	The present disclosure is related to an eco-technological alternative which provides an integral energy assurance, using renewable and alternative energy in hybrid energy systems, applicable to different types of constructions, to solve a new living environment based on the environmentally friendly housing construction, energy-saving equipment installation and the optimization of the existing natural resources. Two photovoltaic systems (PVS) were integrated into the housing, the first one, of 1 kW is interconnected to an inverter to generate AC to power basic appliances. The other 600 W PVS, generates DC; which is used for LED lighting and also for powering the fuel cell controller (CC). The power electronics is controlled by a PLC which helps regulating and distributing energy in the housing. A rainwater recovery system is also integrated in conjunction with a solar heater to complement the sustainability concept. A hydrogen technologic system was also added, to use the surplus energy from the PVS in an electrolyser to generate hydrogen; same that is stored in two metal hydride tanks to feed 500 W of DC and also used as a backup system.	Ecological Bioclimatic System for Supplying Water and Energy in a Housing.	Y02B, Y02E, Y02P, Y04S
493622499	The present invention provides a method and a sensor node for providing an adaptive sampling in a wireless sensor network. According to an embodiment of the present invention, the sensor node providing an adaptive sampling in a wireless sensor network comprises: a sensing part sensing data; a control part processing the sensed data; a communication part transmitting the processed data to the outside according to control of the control part; and an energy storing part storing energy according to the control of the control part, or supplying the energy to at least one among the sensing part, the control part and the communication part. If harvesting quality is more than or equal to a predetermined level, the control part increases a current sampling cycle to control so that the sensing part senses the data, and controls so that the energy storing part stores the energy to be harvested. The energy of the sensor nodes composing the wireless sensor network can be saved, thereby ensuring self-sustainability.	Method and Sensor Node for Providing Adaptive Sampling in Wireless Sensor Networks	Y02D, H04W
497949637	The invention discloses a BP neural network photovoltaic power prediction method based on an information fusion theory. As a new energy source having the advantages of environmental protection, sustainability and short construction period, photovoltaic power generation has become a major force in power generation. However, due to the intermittence, randomness and volatility of photovoltaic power generation, a high ratio of effective access is difficult to guarantee, thus a certain influence on safe operation and dispatching of a power grid is caused. Precise photovoltaic power prediction can effectively solve the problem and accelerate the development of photovoltaic power generation. The BP neural network photovoltaic power prediction method fully considers influence factors of photovoltaic power generation and fuses the influence factors into an influence factor lambda, utilizes the advantage of adjustable structure of a BP neural network to predict moments with great fluctuations precisely, can realize efficient and precise prediction on photovoltaic power, achieves the purposes of peak shaving and valley filling as well as smoothing power output through cooperative use with other energy sources or energy storage systems, improves the superiority of grid-connected operation of a photovoltaic power generation system, and provides a reference basis for dispatching of a power grid.	Bp Neural Network Photovoltaic Power Prediction Method Based on Information Fusion Theory	Y02E, Y04S, G06N, G06Q
334687748	There is provided a greenhouse gas emission trading system including at least one battery for storing electric power, a measuring unit that measures, after a first amount of electric power has been stored in the battery, when a power supply request is received from the outside, a power amount of electric power having been stored in the battery, and a calculation unit that calculates greenhouse gas emissions for notification to the outside, based on measurement results by the measuring unit and information about emissions of greenhouse gases that were emitted when generating the first amount of electric power.	Greenhouse Gas Emission Trading System, Greenhouse Gas Emission Trading Apparatus, Greenhouse Gas Emission	Y02B, Y02P, Y04S, F01K, H02J

		Trading Method and Program	
458082962	The invention provides a high-rise building domestic sewage green purification and reuse system and a control method thereof. The system comprises: a rainwater-sewage power generation unit, a wind-solar complementary power generation unit, a storage battery power supply unit, a sewage disposal-sewage purification unit, a purified water reuse unit and a control unit. The system not only can fully utilize the characteristics of high-rise buildings, comprehensively utilizes wind energy, solar energy, rainwater and domestic sewage's mechanical energy, but also applies the green energy to a domestic sewage purification and reuse system, realizes purification and reuse of high-rise building domestic sewage, and maximumly recovers and utilizes various energy of high-rise buildings. With a simple structure, the system provided by the invention is convenient for transformation and installation of existing high-rise buildings, and also can effectively improve urban environment.	High-Rise Building Domestic Sewage Green Purification and Reuse System and Control Method Thereof	Y02B, Y02E, Y02W, C02F, G05B, H02J, H02S

Source: PATSTAT 2019a.

Table C2: Exemplary patents linked to Cluster 2.

Appln_id	Abstract	Title	CPC subcl.
21523054	A method for storing and removing CO2 from sour bio-gas and alcohol digester biofuel product-propellant mixtures featuring reversible foam-aerosol tri-state containment is described and applied to an autonomous fuel station. The method provides lowered fuel-cycle, odour-free, lower environmental impact storage of VOCs and can source its fuel energy from biomass and various particulate waste streams in foam farms. The liquid and gas fuel products are stored in the first blimp 1 at low pressure in foams over digester pits which are jetted with surfactants to encapsulate VOCs and greenhouse gases and this foam is then reversibly pressurised with bio and petroleum gas-sourced propellants into tri-states and chilled and set in gels. CO2 is removed from biogas containing methane propellant by the differential sorbing action of agitated oil and water-miscible states within the emulsified tri-state. Upon settling, the lower tri-state containing the water-solved fugitive CO2 state is run-off into a second foam storage blimp 2. The upper settled floating tri-state stores the water-fugitive differentially solved oil product with miscible methane propane and butane fuel propellants.	Fuel Store Featuring Removal of CO2	Y02C, Y02E, B01D, C10L, F17C
422975773	A method and a system for managing a greenhouse gas emission quantity focusing on distribution processes are provided. The method for managing a greenhouse gas emission quantity according to an embodiment of the invention measures energy usage quantities at greenhouse gas emission sources for respective greenhouse gas emission sources, and calculates the greenhouse gas emission quantities for business processes, to which the greenhouse gas emission sources are allocated, based on the energy usage quantities. According to the present invention, the greenhouse gas emission quantities are calculated for the respective distribution processes, such that the greenhouse gas emission quantities proper for circulation/distribution companies/industries are allowed to be calculated and managed.	Method and System for Managing Greenhouse Gas Emission Quantity in Logistics Processes	Y02C, Y02P, G06Q

Source: PATSTAT 2019a.

Table C3: Exemplary patents linked to Cluster 3.

Appln_id	Abstract	Title	CPC subcl.
328388261	The claimed subject matter relates to an architecture that can facilitate analysis, processing, or reporting in connection with energy consumption data and/or emissions or sustainability factors associated with an automation process. In particular, the architecture can obtain process-level or machine- or device-level energy consumption data collected during execution of an automation process. The data can be analysed or processed, with general or application-specific results output to a specified recipient and/or formatted (e.g., parsed, filtered, or transformed) according to a general or application-specific scheme.	Real Time Energy Consumption Analysis and Reporting	Y02P, G06Q
500069365	One or more devices, systems and/or methods for managing waste are provided. For example, a waste profile may be generated, based upon inputs received from a first device via a profile interface of a waste data aggregation platform. An approval notice may be received, via an approval interface of the waste data aggregation platform, from a second device. Shipping information may be entered into a plurality of digital forms of a shipping documentation interface based upon the waste profile and/or one or more inputs received from the first device and/or the second device. A shipping order for shipment of the waste may be generated. One or more notifications corresponding to the shipment of the waste may be transmitted to one or more devices associated with the shipment. The waste profile may be analysed to generate a risk report, a sustainability report, a landfill report, a raw material source report, etc.	Waste Management System	Y02W, G06F, G06Q, H04L
273397915	A system for generating standardized greenhouse gas emission reduction credits based on mitigation of greenhouse gas emissions at a site resulting from use of renewable carbon as a fuel at the site to produce energy in substitution for previous use of a fossil fuel at the site, includes a computer, at least one database accessible by the computer and containing past site-specific data and subsequent site-specific data regarding the at least one variable impacting emissions of the greenhouse gas generated from the use of the renewable carbon as a fuel for generating energy at the site, a greenhouse gas emissions modelling module, and at least one interface to the computer, for outputting a report of the identified quantity of the standardized greenhouse gas emission reduction credits.	Technique for Determining and Reporting Reduction in Emissions of Greenhouse Gases at a Site	Y02A, Y02E, Y02P, Y04S, G06Q
6657465	The present invention relates to an all-facade daylighting window and intelligent all-weather solar building, belonging to the field of 'wall change, window change and heat change' system, new energy-saving building and green energy source utilization technology. Said invention includes raking cornick shading-free fence all-facade daylighting curtain wall window, rotating heat curtain tube, shading-free heat storage equipment and intelligent programmable control solar house. Its window daylighting rate can be raised by once, utilization rate of green energy source exceeds the standard of ecological dwelling house and the energy-saving rate is greater than 50%.	Curtain Wall Window Structure and Full Daylight Solar Air-Conditioner	Y02A, Y02B, Y02E

Source: PATSTAT 2019a.

Table C4: Exemplary patents linked to Cluster 4.

Appln_id	Abstract	Title	CPC subcl.
338109443	It is disclosed an energetically-autonomous transportation vehicle characterised by infinite power collection and enough force to overtake other vehicles. It is characterised by perpetual mechanical motion, by utilisation of green energy sources, by dynamic power and mechanical energy generation, by automatic distribution of electromagnetic fields according to needs, by energy recovering what leads to the use of forces inside and outside the system. (unclear source text)	Energetically-Autonomous Transportation Vehicle Using Multiple Green Energy Sources	Y02T, B60L
375788504	The invention discloses a new energy rail bus transit system. The system is provided with a special driving channel, a seal and special bus station and an intelligent bus control system, wherein the special channel is provided with rails and a rail bus set matched with the rails; the rail bus set is wirelessly supplied with power by a vehicle power supply; and a special bus station platform is provided with a ground charging station; With the adoption of the new energy rail bus transit system, the service level and excellent performances, such as comfort, stability and no noise when passengers take the bus, rapidity, safety and intelligent control level, can be compared with those of metros and light rails; the unit construction cost is low; and the comprehensive upgrade and upgrade of a BRT (bus rapid transit) mode is realized in less fund. With the adoption of the new energy rail bus transit system, the wireless power supply by the vehicle power supply is implemented, the advantages of a super capacitor, such as quick charging and discharging ability, long cycle charging life and high power density are given into play; and a solar generator is used for reserving and supplying power for the rail bus set, so that the new energy rail bus transit system has no pollution emission, is obvious in energy-saving effect, and has the prospect of sustainable development.	New Energy Rail Bus Transit System	Y02T

Source: PATSTAT 2019a.

APPENDIX D: THE INDUSTRY CLASSIFICATION BENCHMARK (ICB)

The Industry Classification Benchmark (ICB) allocates companies to subsectors based on the nature of their business. This allocation considers an assessment of revenue source data and other publicly available information. The ICB classification was launched in 2005 and updated in 2019. It distinguishes between 11 Industries, 20 Supersectors, 45 Sectors, and 173 Subsectors. A list of the industries, their definitions, and the sectors that compose them is presented in Table D1.

Table D1: ICB definitions for industries and their composing sectors as defined by FTSE Russel (2021).

Industry	Definition	Sector
Technology	Companies that are primarily engaged in the advancement of the information technology and electronics industries. It includes companies developing integrated computer systems and services, application software not specific to industry market segments, and digital platform providers that generate revenue from advertising contents and derive subscription fees from an advertiser. Also included are companies that develop next generation electronics and related components. Disruptors leveraging "new" technology be placed in the Technology Industry. Rather, individual company technology applications and services will be reviewed as to the markets they serve. Examples include: companies that provide health care, technology equipment, electronic entertainment (video games), e-retailers, and transaction processing service companies.	Software & Computer Services Technology Hardware & Equipment
Telecommunications	Contains companies that own and operate telecommunication infrastructures to provide content delivery services. Also included are manufacturers of telecommunication equipment and components.	Telecommunications Equipment Telecommunications Service Providers
Health Care	Consists of companies that manufacture health care equipment and supplies or that provide health care-related services such as lab services, in-home medical care and operate health care facilities. Also included are companies involved in research, development and production of pharmaceuticals and biotechnology products, and medical marijuana producers.	Health Care Providers Medical Equipment & Services Pharmaceuticals & Biotechnology
Financials	Consists of companies engaged in savings, loans, security investment and related activities such as financial data and information providers. Other examples include mortgage/consumer/corporate financing, investment banking and brokerage, asset management and custody, insurance, and Mortgage REITs.	Banks Finance & Credit Services Investment Banking & Brokerage Services Mortgage Real Estate Investment Trusts Closed End Investments Open End & Miscellaneous Investment Vehicles Life Insurance Non-life Insurance
Real Estate	Consists of companies engaged in real estate investment, development, and other real estate related services. Also includes Equity REITs. Mortgage REITs are classified under Financials Industry.	Real Estate Investment & Services Development Real Estate Investment Trusts
Consumer Discretionary	Contains companies that provide products and services directly to the consumers, and their purchasing habits are non-cyclical in nature (discretionary). Includes companies that manufacture and distribute Household durable goods, apparel, home electronic devices, leisure equipment, and automotive and related parts. The services segment includes hotels, restaurants, retail/e-retail, passenger transportation, and other leisure facilities. Also includes media companies that engage in entertainment content creation and traditional advertisement. Excludes web-portal/hosts that generate revenue through advertisement,	Automobiles & Parts Consumer Services Household Goods & Home Construction Leisure Goods Personal Goods Media Retailers Travel & Leisure

	which are classified under Technology – Consumer Digital Services.	
Consumer Staples	Contains companies that provide products and services directly to the consumers, and their purchasing habits are cyclical in nature (staples). Includes companies that manufacture, distribute, and/or retail food, beverages, and other non-durable household goods. It also includes drug-retailing companies as well as agriculture, fishing, ranching and milling companies.	Beverages Food Producers Tobacco Personal Care, Drug & Grocery Stores
Industrials	Consists of companies engaged in manufacturing and distribution of capital goods and provider of business support services. Includes aerospace, weapons/defense, commercial vehicles, construction materials, industrial machinery and equipment manufacturers. The service segment includes commercial transportation services, business support, maintenance and security services, international trade, transaction processing, and diversified logistic support services.	Construction & Materials Aerospace & Defense Electronic & Electrical Equipment General Industrials Industrial Engineering Industrial Support Services Industrial Transportation
Basic Materials	Consists of companies that extract or process raw materials, and manufacturers of semi-finished goods such as chemicals, textile, paper, forest products and related packaging products. Metals and minerals miners, metal alloy producers, and metal fabricators are also included.	Industrial Materials Industrial Metals & Mining Precious Metals & Mining Chemicals
Energy	Contains of companies that engage in energy extraction, process, and production activities and produce related energy equipment. Includes both renewable and non-renewable energy companies. Companies that primarily engages in distribution of energy are classified in Utilities Industry.	Oil, Gas & Coal Alternative Energy
Utilities	Contains companies that distributes electric, gas, and water. Most companies in this industry are heavily affected by government regulation. Also includes companies that provide waste, recycle, and related environmental services.	Electricity Gas, Water & Multi-utilities Waste & Disposal Services

APPENDIX E: TOP 60 R&D SCOREBOARD COMPANIES BASED ON SPECIALISATION ADVANTAGES

Table E1: Total number of relevant specialisations (Rel. Spec.) from the top 60 scoreboard companies for the considered period (2016-2018).

Short Name	Country	Total Spec.	Irr. Spec.	Rel. Spec.	Share Rel. Spec.	Spec. Cl. 1	Spec. Cl. 2	Spec. Cl. 3	Spec. Cl. 4
Hitachi	Japan	123	93	30	24%	13	2	7	8
General Electric	US	143	119	24	17%	9	2	7	6
Panasonic	Japan	141	117	24	17%	14	1	5	4
Honeywell	US	114	92	22	19%	10	1	7	4
Mitsubishi Motors	Japan	113	92	21	19%	6	1	6	8
Siemens	Germany	90	69	21	23%	12	1	6	2
Mitsubishi Heavy	Japan	147	127	20	14%	8	1	6	5
Toshiba	Japan	104	84	20	19%	13	1	3	3
Nari Technology	China	43	23	20	47%	15	1	3	1
Korea Electric Power	South Korea	78	60	18	23%	9	1	4	4
Mitsubishi Electric	Japan	69	51	18	26%	11	0	3	4
Bae Systems	UK	90	73	17	19%	6	0	5	6
NTN	Japan	60	43	17	28%	8	0	3	6
United Technologies	US	131	114	17	13%	6	1	5	5
Toyota Motor	Japan	120	103	17	14%	8	1	3	5
Boeing	US	132	115	17	13%	7	0	6	4
Denka	Japan	75	58	17	23%	9	0	4	4
Ford Motor	US	91	75	16	18%	9	0	1	6
Robert Bosch	Germany	120	104	16	13%	6	0	5	5
BMW	Germany	104	88	16	15%	7	0	4	5
Toyota Industries	Japan	104	88	16	15%	8	0	3	5
LG Electronics	South Korea	97	81	16	16%	9	0	4	3
LS	South Korea	54	38	16	30%	9	0	4	3
Omron	Japan	71	55	16	23%	10	0	4	2
Hisense Electric	China	72	56	16	22%	10	1	3	2
Kawasaki Heavy Industries	Japan	109	94	15	14%	1	1	2	11
BYD	China	68	53	15	22%	7	0	1	7
Sumitomo Electric	Japan	106	91	15	14%	8	1	0	6
General Motors	US	101	86	15	15%	6	0	4	5
Denso	Japan	96	81	15	16%	10	0	0	5
Delta Electronics (Taiwan)	Taiwan	49	34	15	31%	12	0	1	2
Alphabet	US	60	45	15	25%	9	2	3	1
Inventec	Taiwan	54	39	15	28%	11	0	3	1
Emerson Electric	US	83	68	15	18%	9	0	6	0
Cypress Semiconductor	US	28	13	15	54%	12	0	3	0
NXP Semiconductors	Netherlands	50	35	15	30%	15	0	0	0
Kia Motors	South Korea	96	82	14	15%	7	0	2	5
Hyundai Motor	South Korea	94	80	14	15%	7	0	2	5
Tesla	US	32	18	14	44%	6	0	5	3
Eaton Corporation	Ireland	46	32	14	30%	8	0	4	2
Metallurgical of China	China	98	84	14	14%	6	2	6	0
Schneider	France	31	17	14	45%	11	0	3	0
Honda Motor	Japan	88	75	13	15%	5	1	0	7
Nio	Cayman Islands	29	16	13	45%	9	0	0	4
Volkswagen	Germany	29	16	13	45%	8	0	2	3
ABB	Switzerland	35	22	13	37%	8	0	3	2

Rockwell Automation	US	31	18	13	42%	9	0	2	2
Xiaomi	Cayman Islands	48	35	13	27%	12	0	1	0
Nec	Japan	52	39	13	25%	13	0	0	0
Kubota	Japan	55	43	12	22%	4	0	3	5
Continental	Germany	66	54	12	18%	5	0	2	5
Crrc China	China	61	49	12	20%	6	0	1	5
Raytheon	US	82	70	12	15%	8	0	2	2
Kobe Steel	Japan	115	103	12	10%	4	1	6	1
Illinois Tool Works	Us	107	95	12	11%	9	0	2	1
Johnson Controls	Ireland	39	27	12	31%	9	0	2	1
Ntt	Japan	46	34	12	26%	10	1	0	1
Telefonica	Spain	15	3	12	80%	11	0	1	0
Thales	France	54	42	12	22%	11	1	0	0
Continental	Germany	66	54	12	18%	5	0	2	5

Source: PATSTAT 2021.

Note: The top 50 companies would be selected originally to reflect the same choice shown in the 2021 EU climate neutrality report (e.g., Table 4.1., pg. 56), but there are 10 companies tied in the 50th position with 12 relevant specialisations. Accordingly, it was decided to include all this tied companies in Table E1. Furthermore, “Irr. Spec.” stays for specialisation advantages a company has but that are not relevant to deploying digital sustainability technologies. “Spec. Cl.” Stand for the number of specialisation advantages a firm has in a selected cluster (from cluster 1 to 4).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

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

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science

Publications Office of the European Union