European Commission

Publications as a measure of innovation performance: Selection and assessment of publication indicators

Provision of technical assistance and study to support the development of a composite indicator to track clean-energy innovation performance of EU members

Independent Expert Report

Simon Provençal, Science-Metrix
Paul Khayat, Science-Metrix
David Campbell, Science-Metrix
January 2021

Research and Innovation

Publications as a measure of innovation performance: Selection and assessment of publication indicators

Provision of technical assistance and study to support the development of a composite indicator to track clean-energy innovation performance of EU members

European Commission
Directorate-General for Research and Innovation
Directorate C — Clean Planet
Unit C.1 — Clean Energy Transition
Contact Daniele Poponi
Email daniele.poponi@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in January 2021.

PDF ISBN 978-92-76-38680-3 doi:10.2777/43576 KI-02-21-750-EN-N

Luxembourg: Publications Office of the European Union, 2021

EUROPEAN COMMISSION

Publications as a measure of innovation performance: Selection and assessment of publication indicators

Provision of technical assistance and study to support the development of a composite indicator to track clean-energy innovation performance of EU members

In association with:

Directorate-General for Research and Innovation
Clean Planet

2021

EN

Table of Contents

Abbreviations

CEII	Clean Energy Innovation Index
CNIPA	Chinese National Intellectual Property Administration
EIS	European Innovation Scoreboard
EPO	European Patent Office
EU27	European Union Member States
$HCP_{10\%}$	Highly cited publications among the 10 % most cited
JPO	Japan Patent Office
KA	Key action
KIPO	Korean Intellectual Property Office
MIM	Mission Innovation Member
NSE	Natural sciences and engineering
RC	Relative citation
SCP	Share of publications cited by patents
SI	Specialisation index
SIP	Share of international co-publications
SOA	Share of open access publications
SPP	Share of public/private co-publications
SSH	Social sciences and humanities
STP	Share of transnational co-publications
USPTO	United States Patent and Trademark Office
WEC	Weighted eigenvector centrality
WIPO	World Intellectual Property Organization

1 Introduction

This report is the fourth deliverable of a study supporting the development of the Clean Energy Innovation Index (CEII), which will track the clean energy innovation performance of EU Member States (EU27, without the United Kingdom) and Mission Innovation members (MIM). This index will contribute to the overarching aim of assessing progress in clean energy innovation by analysing output-related indicators. The CEII will be a composite indicator consisting of three core dimensions (i.e. scientific publications, patents and trade), each with multiple indicators capturing various aspects of the innovation system. This report covers the work on using publications as a measure of innovation performance. The other two dimensions are covered in separate reports.

Science-Metrix built nine data sets of publications, one for each of the SET Plan key actions (KA):

- KA1: Performant renewables/Reduced technology costs[1]
- KA2: New technologies & services for consumers
- KA3: Resilience & security of the energy system
- KA4: New materials & technologies for buildings
- KA5: Energy efficiency in industry
- KA6: Competitive in the global battery sector (e-mobility)
- KA7: Renewable fuels
- KA8: Carbon capture, utilisation and storage
- KA9: Nuclear safety

The data sets were built by using an extensive list of keywords specific to the KAs and querying them in the title, abstract and author keywords of the peer-reviewed publications indexed in Scopus, a bibliographic database with a broad coverage of the scientific literature. Using these data sets, Science-Metrix calculated 10 publication-based indicators of potential relevance to the policy context surrounding the development of the CEII. This report focuses on assessing the pertinence and robustness of these indicators for inclusion in the CEII through a qualitative review and quantitative assessment of the data computed by country (EU27 and MIMs) for each KA.

The report is structured as follows. In Section 2, we discuss the main challenges of using publications as a measure of innovation performance and lay out the methodology used to assess the soundness of each indicator for consideration in the CEII. In Section 3, we calculate the indicators, apply the methodology to assess their soundness, discuss the results and determine whether or not they are appropriate for inclusion in the CEII. Section 4 concludes by listing the indicators that were selected.

[1] The data set specific to KA1 was further desegregated into five sub-KAs.

2 Publications as a dimension for measuring innovation performance

Peer-reviewed publications are commonly used to measure research performance, especially in the natural and applied sciences, since they are commonly used to disseminate new knowledge in these areas. Scientific publications, as measured via inclusion in bibliometric databases (e.g. Scopus, Web of Science), mostly consist of three document types: journal articles, published to disseminate the results of novel research; review articles, published to aggregate and summarise the research findings published thus far in a specific topic; and conference papers, published as in-depth reports of presentations given at conferences. Conference papers are particularly relevant to the field of clean energy technologies since conferences are a prominent venue to disseminate research findings in the technological and engineering fields. For example, conference papers account for 66 % of the peer-reviewed scientific output indexed in Scopus for the 2015–2019 period in Information & Communication Technologies; they account for 38 % in Engineering.

Scientific publications also provide a wealth of information on their authors that enables us to extrapolate collaboration practices at various aggregation levels (e.g. between researchers, activity sectors, countries, regions) according to their affiliation. Additionally, bibliographic data on a publication's or a patent's references to the scientific literature make it possible to track which publications, and their corresponding authors, were most credited by their peers for their intellectual contribution to the advancement of science and technology. One of the key beneficial aspects of publication statistics is that they are readily available at a high level of granularity. However, there are also some shortcomings in using publication statistics for measuring research performance. In Section 2.1, we identify publication-based indicators worth considering for inclusion in the CEII and discuss potential shortcomings of using them to monitor research performance. In Section 2.2, we detail the quantitative methodology used to assess the robustness and value of each indicator for inclusion in the CEII.

2.1 Key challenges to publication-based indicators

When dealing with publication data, several issues need to be considered to properly interpret them. Some of these are merely practical challenges that require a transparent and consistent approach, but do not pose challenges beyond that. Others are more fundamental in nature and cause debate around the usefulness of publication data for measuring scientific performance. Some challenges only pertain to specific indicators. As a reminder, the publication-based indicators proposed for this project are as follows:

- Number of publications
- Specialisation index (SI)
- Share of international co-publications (SIP)
- Share of transnational co-publications (STP)
- Weighted eigenvector centrality (WEC) in the world's co-publication network
- Weighted eigenvector centrality in the Member States' co-publication network
- Share of open access publications (SOA)
- Share of public/private co-publications (SPP)
- Share of highly cited publications (those among the 10 % most cited; $HCP_{10\%}$)
- Share of publications cited in patents (SCP)

These indicators were chosen to measure the scientific aspects of core relevance to the policy context surrounding the development of the CEII (i.e. the SET Plan, the Energy Union, and the Mission Innovation initiative): the level of research output, the collaboration of research activities, the openness of research activities and the impact of research activities. For a quick access summary of indicator description, relevance, comparability, availability (and data sources), quality and assessment, the reader is referred to Annex 1.

2.1.1 Data availability

The metadata of scientific publications are indexed in bibliographic databases. For this project, Scopus was used for its broad coverage of the scientific literature, which includes more regional and non-English-language journals than some competing databases, as well as its wide coverage of conference proceedings, which represent a very important dissemination media for scientific discoveries in technological and engineering fields.

A comprehensive bibliometric database indexing all scientific articles published worldwide and indexing all necessary information for producing large-scale bibliometric studies does not currently exist. All databases, including Scopus, have their own biases with regard to geography (some countries are better covered than others), language (English is usually better covered than other languages) and thematic coverage (some fields of science are better covered than others).

Another database worth mentioning is 1findr, which tracks the availability of publications in open access. 1findr's coverage is wider than that of Scopus, but it lacks the necessary metadata to calculate a wide range of bibliometric indicators. The open access status of publications it provides can nevertheless be exploited by linking its content to Scopus.

2.1.2 Practical challenges

Some peculiarities pertaining to bibliometric data need to be considered in order to interpret the indicators properly. For example, the language coverage of bibliometric databases, which is somewhat biased towards the literature written in English, can be a shortcoming. For instance, whereas research questions in the natural sciences and engineering (NSE) tend to be universal, social sciences and humanities (SSH) research subjects are often more local in orientation/focus and, as a result, the target readership is more often limited to a country or region. Consequently, SSH scholars publish somewhat more frequently in a language other than English—and in journals with a national distribution rather than an international distribution—than NSE researchers do. The uninformed or careless use of bibliometrics to benchmark SSH research can thus lead to erroneous conclusions. Notwithstanding that reality, Scopus covers a substantially large number of relevant records for the fields of highest relevance to this study (e.g. Built Environment & Design, Earth & Environmental Sciences, Enabling & Strategic Technologies, Engineering and Information & Communication Technologies), which enables the computation of representative indicators for all countries covered in this study.

Bibliometric databases are also not populated in real time, meaning that it can take up to a year before a publication is indexed. The speed of indexing has historically been slowest for conference papers, which are important to this study. However, this aspect has greatly improved in recent years and Scopus's coverage for a given publication year is estimated to be 98 % complete as early as February of the following year; this means that for publication year 2019, Scopus is 98 % complete as of February of 2020. Of course, there is some variation across countries, but nothing outstanding. Of the 42 countries covered in this study, 60 % have attained the expected figures in terms of number of publications for 2019, and only five have a coverage below 90 % of the expected figures, including Indonesia (72 %), Croatia (85 %), Malta (86 %), Bulgaria (88 %) and Latvia (89 %). The bibliometric indicator most likely to be impacted by such

variation in coverage is the number of publications. All other bibliometric indicators are relative indices that should not be biased with these levels of completeness. Still, special care was paid to the detection of potential outliers in the most recent year of data (i.e. 2019) that could arise from such variation in country coverage. The methodology to identify outliers is detailed in Section 2.2. Based on information currently available, we expect Scopus to provide timely data, as requested in this study (i.e. data from 2015 up to the latest year), for all 42 countries included in constructing the CEII.

2.1.3 Indicator-specific shortcomings

Number of publications

The number of publications is a straightforward indicator, but it does not convey information that can be used to make a direct comparison between countries, since larger economies naturally tend to publish more papers. A normalising variable that takes country size into consideration is therefore needed. The ideal denominator that accomplishes as much is population data, which are more stable than other data, such as GDP. Population data are provided by the World Bank[2]. One shortcoming worth mentioning related to the World Bank's databank is that none of Taiwan's statistics are included, whereas Science-Metrix adds Taiwan's publications to China's total, as well as Hong Kong's and Macao's publications. Hong Kong's and Macao's data are indexed by the World Bank, and were added to China's population for this study, but Taiwan's population is still lacking. Science-Metrix used Taiwan's population data provided by the United Nations[3] and added them to China's total.

As a last note, comparing the number of publications across the KAs is also problematic because there is some variation in precision and recall between the data sets of publications that define the KAs. It is therefore more appropriate to compare the ranking of countries between KAs instead of their raw number of publications, whether or not they are normalised. However, a rescaling of this indicator on a uniform scale applied to all KAs, like the one described in Section 2.2, enables comparison between KAs. Within each KA, however, the volume of publications can be compared between countries without issue.

Share of open access publications

A shortcoming specific to the share of open access publications (SOA) concerns conference papers. In order to measure the extent of the scientific literature available in open access, the coverage of 1findr—the database used to flag the open access status of articles—needs to overlap with Scopus, which contains the necessary metadata to calculate the SOA at country level. However, very few conference papers are indexed in 1findr. As a result, of the 740,283 articles included in the data sets of the KAs (Table 1 farther below), 67.8 % are indexed in 1findr, meaning that the open access status of articles can only be accessed for that share of the data sets. That share still represents a total of 501,838 articles, which is the actual denominator used to compute the SOA, a sufficiently high number to calculate representative statistics on open access. Lastly, to account for the embargo period—a period of time imposed by some publishers before an article can be made available in open access—the SOA was calculated for articles published between 2013 and 2017. However, beyond the embargo effect, the backfilling of older articles made available in open access is continuously ongoing as the open access movement gains traction. The growth of the SOA can therefore be underestimated depending on the version of the database we are using. To mitigate this shortcoming, the

[2] https://data.worldbank.org/indicator/SP.POP.TOTL

[3] https://population.un.org/wpp/Download/Standard/Population/

SOA is normalised by the weighted world average (like many other indicators, explained in more detail in Section 3). Despite this correction, the evolution in the relative placement of countries can nevertheless be affected, for which there is no ideal correction available.

Share of highly cited publications among the 10 % most cited

The share of highly cited publications among the 10 % most cited ($HCP_{10\%}$) is a citation-based indicator. Fundamentally, the value of a citation is at the heart of a constant debate within the bibliometric community and the scientific community more broadly. The assumption is that a citation reflects not only the visibility of a piece of work, but also a certain recognition of the significance of that cited work. In practice, citations can indeed reflect significance and novelty in research, but they are also used simply to give context or to voice a critique. That said, it is fair to assume the citations given for more dubious reasons do not significantly bias the impact metrics computed in large-scale studies such as this one (i.e. at the country level). For example, researchers also have the incentive to cite their own work to boost their own citation score, which potentially creates a conflict of interest. Although researchers citing their own work is not necessarily scientific malpractice and is legitimate in most cases, the share of $HCP_{10\%}$ calculated in this study excludes self-citations as is currently done for the European Innovation Scoreboard (EIS). However, given the scale of the study at which citation-based indicators are calculated (i.e. country level), removing self-citations from these indicators has little influence on the resulting scores. Despite these uncertainties, citations remain an easy-to-use, discrete piece of quantitative information efficiently applicable to broad scale studies to evaluate scientific research. In this respect, citation-based indicators provide rich information on the relative value of different pieces of research that would be onerous to gather through other means.

One shortcoming that concerns citation practices within the scientific literature is the fact that they differ greatly between the different fields of science and document types. For example, a typical article of the health sciences receives far more citations than a typical article of the arts and humanities. Reviews are also cited more frequently than research articles and conference articles. To mitigate this bias, citation counts are normalised by scientific subfield and document type, as well as by publication year, to produce what are known as relative citation (RC) scores. The latter normalisation considers the fact that older articles have had more time to accumulate citations compared to articles published recently. In general, most publications are either cited only a few times or not at all; in other words, the distribution of citations is heavily skewed towards the lower end of the distribution. This applies to all documents but is significantly more pronounced for conference proceedings, given that they receive far less visibility than journal articles. So much so that in some subfields and years, a small number of citations can result in a very large RC score. For example, in 2016, conference papers in the subfield of Nanoscience & Nanotechnology were cited on average 1.7 times whereas journal articles were cited on average 26.6 times. In this example, a conference proceeding with 5 citations would have an RC score of 2.9 while an article with the same number of citations would have a score of 0.19. This significantly distorts the impact of conference proceedings in many subfields and greatly amplifies the impact of documents that may not be that impactful simply owing to the means by which they were published. For this reason, no RC scores are calculated for conference proceedings and they are therefore excluded from the $HCP_{10\%}$. To assess the extent of that exclusion, Table 1 lists the KAs, the number of publications included in each of the data sets that define them and the share of those publications that are conference papers, published between 2015 and 2019. Overall, 31.7 % of all publications specific to the KAs were conference papers, which is significant. The share is above 40 % for a few KAs. However, for all KAs, the number of publications without conference papers, in the many thousands, is sufficient to calculate citation-based indicators accurately.

Table 1: Number of publications and conference papers in each KA (2015-2019)

Key action	Number of publications	Number of conference papers	Share of conference papers
All KAs combined	740,283	234,823	31.7 %
KA1 - Performant renewables/Reduced technology costs	180,924	57,046	31.5 %
Geothermal	8,064	2,612	32.4 %
Hydro	11,030	3,650	33.1 %
Ocean	5,443	2,060	37.8 %
Solar	115,981	31,729	27.4 %
Wind	47,448	19,985	42.1 %
KA2 - New technologies & services for consumers	127,967	59,976	46.9 %
KA3 - Resilience & security of the energy system	195,307	78,567	40.2 %
KA4 - New materials & technologies for buildings	41,487	15,127	36.5 %
KA5 - Energy efficiency in industry	25,507	8,642	33.9 %
KA6 - Competitive in the global battery sector (e-mobility)	78,613	25,988	33.1 %
KA7 - Renewable fuels	129,413	16,025	12.4 %
KA8 - Carbon capture, utilisation and storage	40,524	5,795	14.3 %
KA9 - Nuclear safety	53,538	17,010	31.8 %

Source: These statistics were calculated using data from Scopus (Elsevier)

Another limitation pertaining to the timeliness of data as well as citations concerns the time lag necessary for new publications to accumulate citations from other publications before their scientific impact can be adequately measured. This lag is typically three years (i.e. publication year plus two years). This means that scientific impact metrics can only be computed up to 2017 by using the latest version of Scopus; 2017 being the latest year for which this time lag requirement is met by the database, with citations from other papers in 2017, 2018 and 2019 being fully captured. To ensure consistency in the length of the period (i.e. five years) for which data are provided across all bibliometric indicators, an exception to the reference period (i.e. 2015–2019) was made for the share of $HCP_{10\%}$ (i.e. 2013–2017).

One last shortcoming worth mentioning is the fact that, by definition, the share of $HCP_{10\%}$ includes a small share of publications, which can cause important year-to-year fluctuations at country level that might not capture true performance. For that reason, the share of $HCP_{10\%}$ was calculated for countries that have published at least 30 publications (with an RC score) in any given year.

Share of publications cited in patents

As a citation-based indicator, the share of publications cited in patents (SCP) was treated differently since very few publications are ever cited by patents. Instead of counting the number of citations, the number of articles cited at least once by a patent was counted. The speed of uptake of the scientific literature within the patent literature is slower than within the scientific literature itself because of the time needed to develop inventions and subsequently apply for patents; here, a minimum lag of five years (i.e. publication year plus four years) was applied. Therefore, to maintain a five-year period across all indicators, the SCP was calculated between 2011 and 2015. Conference articles were not excluded from the SCP since they are cited relatively often by patents (4.3 % of clean energy articles published between 2011 and 2015 and cited by patents were journals articles or reviews, while 1.1 % were conference articles). The technological topics discussed at many conferences, such as those of clean energy, align well with patenting activities. A minimum of 30 publications was also applied to this indicator to mitigate yearly fluctuations.

The PATSTAT database was used to link the references of patents to Scopus. PATSTAT includes information on patents from a number of patenting offices around the world,

which all have their own standards and practices. Since there is no universal patenting office, and considering the fact that this study is quite broad in terms of the countries covered, relying on a single national or regional patent office would be problematic since this would create a "home advantage" in favour of one or a few countries, which are naturally inclined to file patent applications at their own national patent office. That said, some markets are so important that some countries patent more inventions in these key markets than in their own. Therefore, selecting a few major patenting offices is sufficient to cover a significant share of the global market of patented inventions, especially high-value inventions. The offices selected for this study were the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), the Chinese National Intellectual Property Administration (CNIPA), the Japan Patent Office (JPO) and the Korean Intellectual Property Office (KIPO). Patent applications filed through the World Intellectual Property Organization (WIPO) are also included; however, the WIPO simply facilitates the application process to multiple jurisdictions and is not a patenting office itself.

As can be seen in Table 2, the USPTO and the EPO represent the lion's share of cited articles in the KA data sets. Only 2,588 articles, less than 10 % of all cited articles, were cited by a patent filed at the CNIPA, the JPO or the KIPO, but not at the USPTO or the EPO, and 1,174 articles were cited by a patent filed through the WIPO but not at any of these five patenting offices.

Table 2: Number of publications in all KAs combined cited by patents according to the patent office (2011-2015)

Description	Number of publications	Share
Number of publications	521,400	—
… cited by at least one patent filed at the USPTO	19,448	3.7 %
… cited by at least one patent filed at the EPO	10,467	2.0 %
… cited by at least one patent filed at the CNIPA	2,263	0.4 %
… cited by at least one patent filed at the JPO	1,525	0.3 %
… cited by at least one patent filed at the KIPO	1,042	0.2 %
… cited by at least one patent filed at the USPTO or EPO	24,761	4.7 %
… cited by at least one patent filed at the CNIPA, JPO or KIPO	4,595	0.9 %
… cited by at least one patent filed at any one of these five offices	27,349	5.2 %
… cited by at least one patent filed through the WIPO	11,242	2.2 %
… cited by at least one patent filed at any one of these five offices or through the WIPO	28,523	5.5 %

Source: These statistics were calculated using data from Scopus (Elsevier) and PATSTAT

Other indicators

The other indicators are not subject to specific shortcomings beyond those described in Sections 2.1.1 and 2.1.2. As a last point, conference papers tend to be less collaborative in practice, which is reflected in the SIP and the STP. However, this is not a shortcoming but rather an artefact of the subject at hand.

2.2 Assessing the soundness of the indicators for their inclusion in the CEII

To ensure the quality of the data across all countries and KAs, two time-series consistency tests were first applied to all indicators, for each country and KA individually. The first test identified potential non-sampling errors (e.g. processing errors such as harmonising errors, database errors, wrong denominator in normalising an indicator) that could lead to inaccurate data points. To detect aberrations in the time series, an automated test for detecting potential outliers was applied to the time series of each country for each indicator. A linear regression model was fitted on the time series of each country. Subsequently, a statistical procedure was applied to test the null hypothesis that

the studentised residual of each data point could have been generated by the fitted model; when the *p*-value of a test was smaller than 0.05, the hypothesis was rejected, implying that the data point was potentially an outlier. Subsequently, these potential outliers were inspected to assess the degree to which they may represent real variations; in other words, the potential outliers automatically identified using the above statistical test were validated manually to differentiate real outliers (bad data or incorrect definition) from false outliers (data points likely representing real fluctuations due, for example, to economic shocks). This exercise is very complex, as some outliers may naturally arise as a result of unknown conditions. Therefore, actions were taken on data points for which there was no ambiguity regarding their status. When outliers were identified, the data sources used to compute the indicator were analysed to detect where the aberrant values might have come from. Subsequently, the problematic data points were flagged.

The second test identified breaks in time series and other possible outliers. A stepwise analysis was conducted via a script that compared the difference between two successive data points to the average difference of the same points before and after, which is defined as the expected value. This analysis highlighted undocumented breaks in time series or changes of regime that were not detected with the first test. Again, when the test failed a defined quality threshold, a manual validation was applied to each point before flagging any problematic ones.

The tests were applied not only to the indicators but to all sources of data that fed into each indicator. For example, for the number of publications per capita (Section 3.1), the tests were applied to the number of publications and to population data. Also, a longer time series (2010 to 2019) than the one used in building the CEII (e.g. 2015 to 2019) was used to enhance the capability of the implemented statistical procedures that detect problematic data points.

It is important to mention that outliers and breaks in time series are not necessarily inaccurate data points. In fact, all indicators in this study were calculated based on publication data, which are generally continuous through time. However, the publication output of smaller countries is much more prone to annual fluctuations. This was observed for some countries and for some KAs, whose fluctuating trends were much more likely to fail the outlier and break in time series tests. Such fluctuations are to be expected and should be interpreted as within an acceptable range to accurately depict the situation in those countries. In those cases, the small output was sufficient justification to ignore these outliers and breaks in time series. Ultimately, no data were deemed exaggerated to the point that warranted a correction, even for countries whose publications from 2019 are still not fully indexed in Scopus. An example of the application of these tests are shown in Section 3.1.

Once outliers and breaks in time series had been investigated, and quality of the data had been assessed, the indicators were transformed to be properly included in the CEII. The bibliometric indicators listed in Section 2.1 measure different aspects of the scientific literature and are scaled differently. Also, the frequency distribution of bibliometric indicators tends to be skewed toward average and below-average countries, while a few countries obtain outstanding scores. In such cases, a transformation (either the natural logarithm plus one or the square root) was applied to the scores to bring the distribution closer to a Gaussian distribution. However, some indicators that do exhibit a Gaussian distribution were not transformed. Finally, all scores were standardised between 0 and 1, based on either the original (not transformed) or the transformed score (if the transformation was applied), to enable comparison across KAs and countries by applying this formula:

$$\frac{S_X - S_{min}}{S_{max} - S_{min}}$$

Where:

S_X Score of country *X*

S_{min} Minimum score among the distribution of countries (all EU27 and MIMs)

S_{max} Maximum score among the distribution of countries (all EU27 and MIMs)

This formula transforms all indicators to a comparable scale (i.e. between 0, the worst performance, and 1, the best performance) in order to aggregate them into the CEII, which contributes to a uniform contribution of all indicators to the CEII.

Section 3 includes a dashboard that lists all the EU27 members and MIMs and their respective score for each indicator. The dashboard includes the original (not transformed) score, the score normalised by the world weighted average (when appropriate), the transformed score (when appropriate) and the standardised score (between 0 and 1). The dashboard also includes each country's global GDP share, based on purchasing power parity in constant 2017 international $ provided by the World Bank[4], to give a sense of the distribution of economy sizes falling above or below the unweighted mean across all EU27 members and MIMs. The countries are clustered according to the unweighted mean and standard deviation of the original (not transformed) or the transformed score (if the transformation was applied) of all EU27 members and MIMs as follows:

- Cluster 1 includes the countries whose score is above the mean plus one standard deviation.
- Cluster 2 includes the countries whose score is above the mean but below the mean plus one standard deviation.
- Cluster 3 includes the countries whose score is below the mean but above the mean minus one standard deviation.
- Cluster 4 includes the countries whose score is below the mean minus one standard deviation.

Assuming a perfectly normal distribution, Clusters 1 and 4 should include 16 % of countries (or 16 % of the world economy) each, and Clusters 2 and 3, 34 % of countries each. However, EU27 members and MIMs represent 75.5 % of the worldwide GDP, meaning that 24.5 % of the world economy is not represented in the dashboards. Since GDP shares are based on the worldwide total, this should be taken into consideration when analysing the share of GDP of each cluster. China's GDP also lacks Taiwan's since, as mentioned previously, the World Bank does not index any statistics for Taiwan.

As a last step, the correlation between all publication indicators was calculated. If two or more indicators are shown to correlate strongly with each other, this is an indication that some indicators are redundant and do not add much value to the CEII. For example, countries with a large share of their publications co-authored in collaboration with international partners tend to be highly impactful. Following this analysis, we recommended excluding some indicators from the CEII for being redundant. This correlation analysis is presented in the Conclusion (Section 4).

[4] https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.KD

3 Assessment of the potential indicators for inclusion in the CEII

This section presents the performance scores for each EU27 member and MIM, and for each indicator listed in Section 2.1. It also documents the results of the outlier and break in series tests. Finally, it analyses the potential of each indicator to be included in the CEII, which is further nuanced by the correlation structure among them in the Conclusion (Section 4).

3.1 Number of publications

Box 3-1: Conclusion regarding the selection of number of publications for inclusion in the CEII

> Following these analyses, we conclude that the number of publications is a pertinent indicator to measure the level of research output in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII, if it is normalised per capita. There are no missing data in either the publication or population data. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

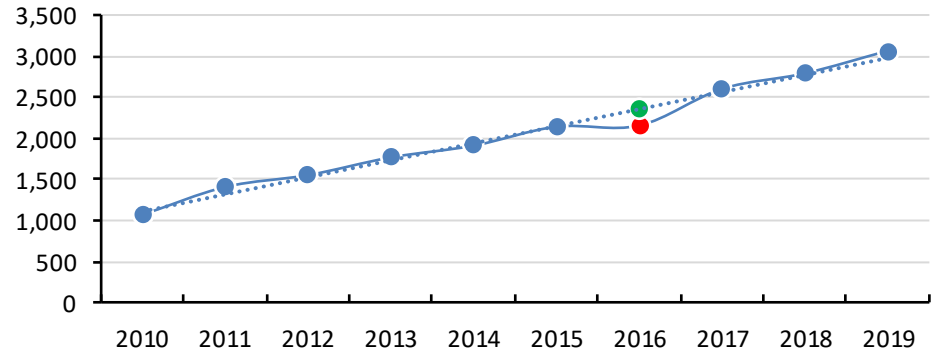
The first bibliometric indicator for consideration in the CEII is the number of publications in fractional counting, which is shown for each EU27 member and MIM in Table 3 (further below) for all KAs combined. The table shows the score for the latest year available (2019 in this case). As mentioned in Section 2.1, for comparison purposes, this indicator is flawed for the simple reason that larger countries naturally tend to publish more articles. Population data were used as a normalising variable that takes the size of countries into consideration.

The outlier and break in time series tests were then applied on the number of publications. For all KAs combined, the test identified one potential outlier at the 95 % confidence level: Brazil's number of publications in 2016. It is worth mentioning that this potential outlier was deemed insignificant at the 99 % level. Brazil's publication output is shown in

Source: These statistics were calculated using data from Scopus (Elsevier)

with the potential outlier in red, which does appear to slightly diverge away from the overall trend. A trendline that best fits the trend is drawn through the series (the dotted line, which is linear in this particular case) and a potential correction point was added (the green dot). The correction is less than 10 % away from the initial value, which would ultimately have a minor if not negligible impact on the indicator. It was therefore decided to leave this potential outlier as is without applying a correction. The number of publications in each KA is much lower compared to the total of all KAs combined and resulted in many countries for many KAs with no clear temporal trend. The outlier test was applied, regardless, to all countries and KAs and identified a few problematic data points. After investigating each of them individually by using a similar strategy, none of them were deemed worthy of a correction. Finally, the break in time series test was applied to all countries and KAs and, after investigating potential problems, none were considered problematic. The tests were applied to the final indicator and identified a few potential outliers, which ultimately did not raise any serious concern either.

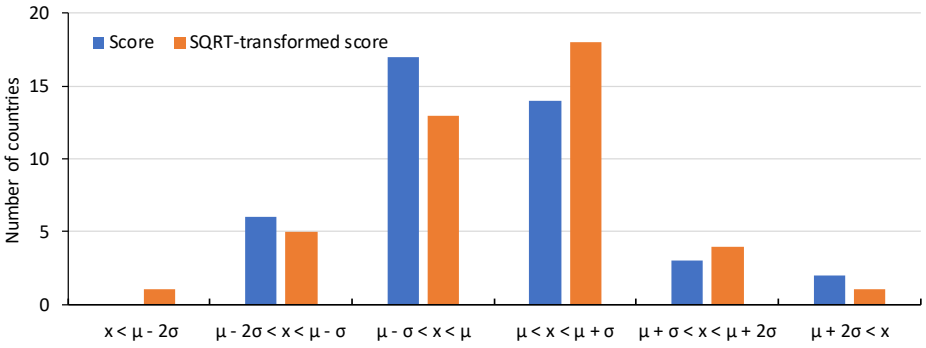
Figure 1: Brazil's publication output (using a fractional count) for all KAs combined, 2010-2019



Source: These statistics were calculated using data from Scopus (Elsevier)

Once the tests had been run on all the data, the number of publications per 1,000,000 population was calculated and is included in Table 3 (the "score" column). As mentioned in Section 2.2, bibliometric indicators tend to be exponentially distributed, with a handful of entities receiving outstanding scores and the bulk of entities receiving average or below-average scores. Once normalised by population, the number of publications was much less skewed (skewness of +0.8, see blue bars in Figure 2) than its original form (skewness of +5.2). Still, the figure shows that a positive skew remains. In order to reduce the influence of countries that perform outstandingly well when constructing the CEII (this could lead to some indicators weighting more heavily in the CEII), some transformations were tested and the square root transformation appeared to best fit a normal distribution, albeit not perfectly. The square root-transformed score is included in Table 3, the frequency distribution of which is also shown in Figure 2 (the orange bars, skewness of -0.1). Table 3 also shows the compound annual growth rate (CAGR) calculated between the first year (2015) and the last year (2019), and the CAGR difference between each country and the EU27 and the MIMs, all of which are calculated based on the original (not transformed) score.

Figure 2: Distribution of countries according to the mean and standard deviation of the number of publications per 1,000,000 population and the square root-transformed equivalent (2019)



Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Table 3: Number of publications (in fractional counting) per 1,000,000 population for all KAs combined

Country	Code	Share in GDP (2019)	Number of pubs. (2019)	Score (2019)	SQRT score (2019)	Score 0-1 (2019)	CAGR		
							2015 to 2019	Gap to EU27	Gap to MIM
World	**World**	**100%**	**174,319**	**23**	**4.77**	**0.28**	**8.4%**	**6.0%**	**0.3%**
EU27	**EU27**	**15.3%**	**29,868**	**67**	**8.17**	**0.55**	**2.4%**	**0.0%**	**-5.7%**
MIM	**MIM**	**69.3%**	**134,450**	**31**	**5.56**	**0.34**	**8.1%**	**5.7%**	**0.0%**
Cluster 1		**2.8%**							
Denmark	DK	0.3%	1,120	193	13.88	1.00	1.8%	-0.6%	-6.3%
Norway	NO	0.3%	841	157	12.54	0.89	6.7%	4.3%	-1.4%
Finland	FI	0.2%	758	137	11.72	0.83	0.0%	-2.4%	-8.1%
Rep. of Korea	KR	1.7%	7,043	136	11.67	0.83	4.9%	2.5%	-3.2%
Sweden	SE	0.4%	1,314	128	11.30	0.80	0.4%	-2.0%	-7.7%
Cluster 2		**14.7%**							
Portugal	PT	0.3%	1,130	110	10.49	0.73	5.6%	3.2%	-2.5%
Cyprus	CY	0.0%	131	109	10.45	0.73	16.1%	13.7%	8.0%
Australia	AU	1.0%	2,748	108	10.41	0.73	4.3%	1.9%	-3.8%
Slovenia	SI	0.1%	224	107	10.36	0.72	7.5%	5.1%	-0.6%
Estonia	EE	0.0%	135	101	10.07	0.70	3.1%	0.7%	-5.0%
Ireland	IE	0.3%	448	91	9.52	0.65	4.3%	1.9%	-3.8%
Luxembourg	LU	0.1%	56	90	9.49	0.65	4.8%	2.4%	-3.3%
Latvia	LV	0.0%	163	85	9.23	0.63	-4.3%	-6.7%	-12.4%
Canada	CA	1.4%	3,069	82	9.04	0.62	2.6%	0.2%	-5.5%
Czech Republic	CZ	0.3%	814	76	8.73	0.59	4.4%	2.0%	-3.7%
Italy	IT	2.0%	4,513	75	8.65	0.59	2.6%	0.2%	-5.5%
Greece	EL	0.2%	798	74	8.63	0.58	6.8%	4.4%	-1.3%
Austria	AT	0.4%	658	74	8.61	0.58	1.4%	-1.0%	-6.7%
United Kingdom	UK	2.4%	4,926	74	8.58	0.58	2.8%	0.4%	-5.3%
Germany	DE	3.4%	5,650	68	8.24	0.55	-0.4%	-2.8%	-8.5%
Netherlands	NL	0.8%	1,170	67	8.22	0.55	0.8%	-1.6%	-7.3%
Spain	ES	1.5%	3,090	66	8.10	0.54	3.5%	1.1%	-4.6%
Belgium	BE	0.5%	749	65	8.08	0.54	-2.3%	-4.7%	-10.4%
Cluster 3		**43.5%**							
Slovakia	SK	0.1%	331	61	7.78	0.52	7.9%	5.5%	-0.2%
Romania	RO	0.4%	1,074	55	7.45	0.49	12.9%	10.5%	4.8%
Croatia	HR	0.1%	223	55	7.40	0.49	7.1%	4.7%	-1.0%
Lithuania	LT	0.1%	146	52	7.24	0.47	1.1%	-1.3%	-7.0%
United States	US	15.8%	16,044	49	6.99	0.45	-1.5%	-3.9%	-9.6%
United Arab Emirates	AE	0.5%	476	49	6.98	0.45	13.1%	10.7%	5.0%
Poland	PL	1.0%	1,702	45	6.70	0.43	9.7%	7.3%	1.6%
France	FR	2.4%	2,965	44	6.65	0.43	-1.8%	-4.2%	-9.9%
Malta	MT	0.0%	21	41	6.40	0.41	10.2%	7.8%	2.1%
Japan	JP	4.0%	4,987	39	6.28	0.40	-1.0%	-3.4%	-9.1%
China	CN	17.7%	55,189	39	6.21	0.39	15.3%	12.9%	7.2%
Saudi Arabia	SA	1.2%	1,056	31	5.55	0.34	15.9%	13.5%	7.8%
Bulgaria	BG	0.1%	211	30	5.51	0.34	25.3%	22.9%	17.2%
Cluster 4		**14.5%**							
Hungary	HU	0.2%	276	28	5.31	0.32	7.2%	4.8%	-0.9%
Chile	CL	0.4%	291	15	3.92	0.21	7.2%	4.8%	-0.9%
Brazil	BR	2.4%	3,075	15	3.82	0.20	8.5%	6.1%	0.4%
India	IN	7.1%	14,464	11	3.25	0.16	19.5%	17.1%	11.4%
Indonesia	ID	2.5%	2,267	8	2.89	0.13	51.4%	49.0%	43.3%
Mexico	MX	1.9%	997	8	2.80	0.12	10.5%	8.1%	2.4%

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Looking at the results in Table 3 more closely, globally 174,319 articles related to all KAs combined were published in 2019; EU27 members contributed 29,868 articles (17.1 %) to that number, and the MIMs, 134,450 articles (77.1 %). At the country level, China published the most articles (55,189 publications), followed at some distance by the United States (16,044 publications) and India (14,464 publications). When considering

population, 23 articles per 1,000,000 population were published worldwide; both EU27 members and MIMs performed better, with 67 and 31 articles per 1,000,000 population, respectively. China is in 34th position, despite its large publication output. The countries that performed the best, in Cluster 1, are all small economies (except for South Korea in 4th position), all of which published more than 125 articles per 1,000,000, population. There is indeed a negative but weak correlation between the score and the GDP share. Denmark, the country that performed the best, published 193 articles per 1,000,000 population.

Globally, the number of publications per 1,000,000 population increased at an annual rate of 8.4 % between 2015 and 2019, mostly driven by MIMs (annual increase of 8.1 %), particularly China (15.3 %), Saudi Arabia (15.9 %). The scores of Indonesia, India, Mexico and Brazil also increased substantially (by 51.4 %, 19.5 %, 10.5 % and 8.5 % annually, respectively), but these countries still remained in Cluster 4, indicating that their output per capita was quite low to begin with. The EU27's score during that time frame slightly increased, by 2.4 % annually. The EU27's two largest economies (Germany and France) saw their score decrease slightly during that time.

Table 4 shows the number of publications per 1,000,000 population for each KA. The number of publications is fractioned between countries but also between KAs, meaning that the sum across the KAs equals the total ("all KAs"). The same countries tend to rank highly across all KAs, while the opposite is true for lower ranking countries. One exception appears to be KA9 (nuclear safety). France and Japan, countries with an important nuclear power capacity, also rank in the top half of the distribution of KA9, while ranking in the bottom half in all other KAs. The global number of publications per capita increased between 2015 and 2019 in all KAs, except KA9, which is also the case for MIMs. The EU27's performance increased in all KAs, except KA1, KA7 and KA9. The reader is referred to the accompanying databook to this report for complete data on trends.

Table 4: Number of publications (in fractional counting) per 1,000,000 population for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**23**	**4**	**4**	**5**	**1**	**1**	**2**	**4**	**1**	**1**
EU27	**EU27**	**67**	**13**	**12**	**11**	**5**	**3**	**5**	**10**	**3**	**5**
MIM	**MIM**	**31**	**6**	**5**	**7**	**1**	**1**	**3**	**5**	**2**	**2**
Denmark	DK	193	55	19	47	15	7	11	30	6	3
Norway	NO	157	47	20	20	15	6	7	16	24	2
Finland	FI	137	18	32	22	9	9	9	21	4	14
Rep. of Korea	KR	136	23	29	23	5	3	15	20	7	11
Sweden	SE	128	23	20	17	11	7	11	23	6	10
Portugal	PT	110	24	22	17	11	7	5	17	4	3
Cyprus	CY	109	27	25	24	13	5	3	8	3	0
Australia	AU	108	22	18	22	8	3	8	17	9	1
Slovenia	SI	107	17	18	16	13	6	8	14	3	11
Estonia	EE	101	16	22	22	17	4	3	14	1	2
Ireland	IE	91	20	27	16	6	4	3	11	3	0
Luxembourg	LU	90	19	36	11	7	4	8	5	0	1
Latvia	LV	85	10	14	16	12	5	7	18	2	1
Canada	CA	82	13	14	17	5	2	8	13	6	4
Czech Republic	CZ	76	8	12	11	12	2	5	13	2	10
Italy	IT	75	11	15	13	7	4	6	11	3	5
Greece	EL	74	14	21	14	6	2	4	10	2	1
Austria	AT	74	9	17	13	7	6	6	12	2	3
United Kingdom	UK	74	16	13	12	5	3	6	10	5	5
Germany	DE	68	15	11	11	2	4	8	9	3	5
Netherlands	NL	67	16	11	11	5	2	4	11	4	3
Spain	ES	66	16	11	11	5	3	4	10	3	3
Belgium	BE	65	13	14	11	5	2	6	6	4	5
Slovakia	SK	61	8	12	8	9	5	6	7	1	3
Romania	RO	55	13	10	10	5	3	5	6	1	2
Croatia	HR	55	12	9	14	3	3	4	7	1	1
Lithuania	LT	52	9	11	5	4	2	2	13	1	5
United States	US	49	8	8	9	2	1	5	7	3	5
United Arab Emirates	AE	49	9	11	9	5	2	3	5	3	1
Poland	PL	45	7	5	9	4	2	3	9	2	2
France	FR	44	8	7	7	2	1	3	6	2	7
Malta	MT	41	20	2	14	1	0	1	0	0	3
Japan	JP	39	8	6	6	1	1	4	7	2	6
China	CN	39	7	5	10	1	1	5	6	2	2
Saudi Arabia	SA	31	8	8	4	1	1	1	4	2	0
Bulgaria	BG	30	5	8	5	1	1	4	4	1	1
Hungary	HU	28	4	5	3	3	2	2	4	2	3
Chile	CL	15	4	2	4	1	1	1	2	0	0
Brazil	BR	15	2	2	3	0	1	0	4	1	0
India	IN	11	2	3	2	0	0	1	2	0	0
Indonesia	ID	8	2	1	1	0	0	0	2	0	0
Mexico	MX	8	2	1	1	0	0	0	2	0	0

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

3.2 Specialisation index

Box 3-2: Conclusion regarding the specialisation index for inclusion in the CEII

> Following these analyses, we conclude that the SI is a pertinent indicator to measure the level of research output in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII. The SI is a normalised indicator that enables comparison of the research intensity of EU27 members and MIMs in clean energy research relative to the world, among themselves and between KAs. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

The specialisation index (SI) is by definition a normalised indicator of a country's research intensity in a given field (clean energy, in this case) weighted by the total publication output at the national and the global levels. Therefore, it can be directly used to compare countries between themselves. Table 5 shows the SI for 2019 and each EU27 member and MIM, for all KAs combined, as well the log-transformed and standardised (on a 0-to-1 scale) equivalents. It is worth mentioning that specialisation is a zero-sum game, meaning that a country cannot be specialised in all fields of research. Similarly, all countries cannot be specialised in the same field. Specialisation in any given field comes at the expense of other fields. That being said, a minority of the EU27 members and MIMs, 15 to be exact, were specialised in clean energy research (with a score above 1), meaning that larger shares of their publication outputs were in clean energy compared to what was observed at the world level. The EU27 as a whole was not specialised in clean energy research (score of 0.85). Its score also decreased by 2.7 % annually between 2015 and 2019. The EU27 members are in fact more specialised in fields such as psychology and cognitive sciences, economics and business, and clinical medicine, at the expense of other fields, such as information and communication technologies, and engineering. However, the most specialised country in clean energy research among all countries (EU members and non-EU members of MI) for which data was collected is Latvia, an EU27 member. It remained at the top in 2019 even though its SI had decreased by 8.3 % annually since 2015, the second strongest decline. The MIMs collectively were more specialised in clean energy research than the EU27, but not significantly more that the world level. Five MIMs were nevertheless specialised enough to figure in Cluster 1. Three of these MIMs (South Korea, India and China) had the largest, 3rd and 4th largest publication outputs in the field (Table 3), weighting significantly more than the other MIMs in their collective SI. To balance things out, the United States, which produced the 2nd largest number of publications in the field, was near last in specialisation. Like the EU27 members, the United States is specialised instead in the fields of the health sciences. The MIMs' SI changed little between 2015 and 2019.

Table 5: Specialisation index for all KAs combined

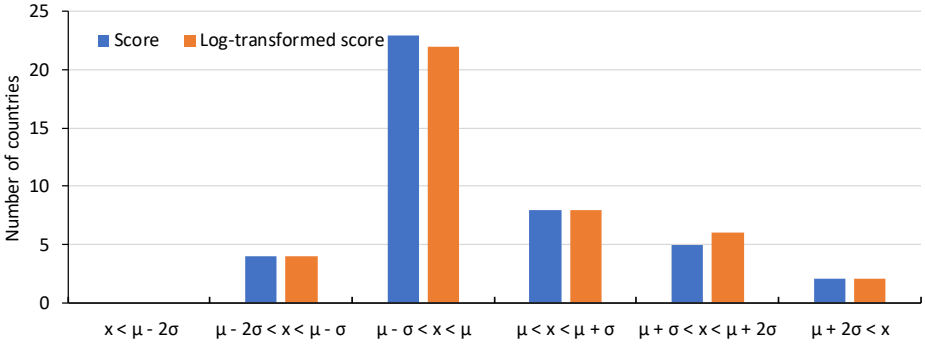
Country	Code	Share in GDP (2019)	Score (2019)	Log score (2019)	Score 0-1 (2019)	CAGR		
						2015 to 2019	Gap to EU27	Gap to MIM
World	**World**	**100%**	**1.00**	**0.69**	**0.30**	**0.0%**	**2.7%**	**0.1%**
EU27	**EU27**	**15.3%**	**0.85**	**0.62**	**0.21**	**-2.7%**	**0.0%**	**-2.6%**
MIM	**MIM**	**69.3%**	**1.03**	**0.71**	**0.31**	**-0.1%**	**2.6%**	**0.0%**
Cluster 1		**28.8%**						
Latvia	LV	0.0%	1.87	1.05	0.72	-8.3%	-5.6%	-8.2%
United Arab Emirates	AE	0.5%	1.77	1.02	0.68	-5.3%	-2.6%	-5.2%
Rep. of Korea	KR	1.7%	1.64	0.97	0.62	-2.1%	0.6%	-2.0%
India	IN	7.1%	1.59	0.95	0.60	5.2%	7.9%	5.3%
Romania	RO	0.4%	1.54	0.93	0.57	7.7%	10.4%	7.8%
China	CN	17.7%	1.44	0.89	0.53	1.2%	3.9%	1.3%
Cyprus	CY	0.0%	1.44	0.89	0.53	1.9%	4.6%	2.0%
Saudi Arabia	SA	1.2%	1.31	0.84	0.47	2.6%	5.3%	2.7%
Cluster 2		**3.8%**						
Estonia	EE	0.0%	1.21	0.79	0.41	-2.5%	0.2%	-2.4%
Denmark	DK	0.3%	1.17	0.77	0.39	-3.6%	-0.9%	-3.5%
Greece	EL	0.2%	1.09	0.74	0.34	0.9%	3.6%	1.0%
Portugal	PT	0.3%	1.07	0.73	0.34	-2.4%	0.3%	-2.3%
Indonesia	ID	2.5%	1.02	0.70	0.31	-11.9%	-9.2%	-11.8%
Finland	FI	0.2%	1.02	0.70	0.31	-5.1%	-2.4%	-5.0%
Norway	NO	0.3%	1.01	0.70	0.30	-2.2%	0.5%	-2.1%
Slovenia	SI	0.1%	0.97	0.68	0.28	4.3%	7.0%	4.4%
Cluster 3		**25.8%**						
Slovakia	SK	0.1%	0.96	0.68	0.27	1.6%	4.3%	1.7%
Lithuania	LT	0.1%	0.95	0.67	0.26	-4.4%	-1.7%	-4.3%
Sweden	SE	0.4%	0.94	0.66	0.26	-3.9%	-1.2%	-3.8%
Italy	IT	2.0%	0.93	0.66	0.25	-4.2%	-1.5%	-4.1%
Luxembourg	LU	0.1%	0.92	0.65	0.25	-1.2%	1.5%	-1.1%
Bulgaria	BG	0.1%	0.89	0.64	0.23	7.2%	9.9%	7.3%
Mexico	MX	1.9%	0.87	0.63	0.22	1.0%	3.7%	1.1%
Ireland	IE	0.3%	0.85	0.62	0.21	-3.4%	-0.7%	-3.3%
Germany	DE	3.4%	0.84	0.61	0.20	-4.6%	-1.9%	-4.5%
Japan	JP	4.0%	0.83	0.60	0.19	-5.0%	-2.3%	-4.9%
Czech Republic	CZ	0.3%	0.82	0.60	0.18	1.3%	4.0%	1.4%
Croatia	HR	0.1%	0.81	0.59	0.18	2.4%	5.1%	2.5%
Spain	ES	1.5%	0.79	0.58	0.16	-2.9%	-0.2%	-2.8%
Austria	AT	0.4%	0.78	0.58	0.16	-3.1%	-0.4%	-3.0%
Brazil	BR	2.4%	0.78	0.58	0.16	0.1%	2.8%	0.2%
Poland	PL	1.0%	0.77	0.57	0.15	3.1%	5.8%	3.2%
Canada	CA	1.4%	0.75	0.56	0.14	-1.9%	0.8%	-1.8%
Malta	MT	0.0%	0.74	0.55	0.13	0.9%	3.6%	1.0%
France	FR	2.4%	0.72	0.54	0.12	-3.6%	-0.9%	-3.5%
Australia	AU	1.0%	0.71	0.54	0.11	-1.4%	1.3%	-1.3%
United Kingdom	UK	2.4%	0.71	0.53	0.11	-2.3%	0.4%	-2.2%
Belgium	BE	0.5%	0.71	0.53	0.11	-5.7%	-3.0%	-5.6%
Cluster 4		**17.2%**						
Hungary	HU	0.2%	0.64	0.50	0.06	1.5%	4.2%	1.6%
United States	US	15.8%	0.56	0.45	0.01	-5.8%	-3.1%	-5.7%
Chile	CL	0.4%	0.56	0.45	0.01	-1.5%	1.2%	-1.4%
Netherlands	NL	0.8%	0.56	0.44	0.00	-4.2%	-1.5%	-4.1%

Source: Statistics calculated using data from Scopus (Elsevier) and the World Bank.

The SI is calculated based on the number of fractional publications in clean energy, as well as the total number of fractional publications in Scopus. Fractional counting is also used to calculate the number of publications per capita. The outlier and break in time series tests were already applied to the number of fractional publications during the process described in Section 3.1; they were therefore not repeated here. The tests were applied to the total number of fractional publications in Scopus, which was substantial for all countries and as a result progressed smoothly. A few potential outliers were identified, but following a visual inspection, none of them seemed to diverge significantly away from their trend. Finally, the tests were applied to the SI itself. The outlier test identified no potential outlier for any country for all the KAs combined. It did identify a few potential outliers for some countries in some KAs. After ignoring those with a small publication output, those that remained were inspected and it was decided to leave these data points unchanged. The break in time series test identified some potentially problematic trends, none of which were deemed worthy of a correction.

Like the number of publications per capita, the SI is also exponentially distributed, meaning that most EU27 members and MIMs (27) were less specialised that the unweighted average (Figure 3). Applying the logarithm to the SI does flatten the distribution, but it is still skewed toward lower ranking countries. The log-transformed score was nevertheless used to standardise the score on a 0-to-1 scale, which is shown in Table 5.

Figure 3: Distribution of countries according to the mean and standard deviation of the specialisation index and the log-transformed equivalent (2019)



Source: Statistics calculated using data from Scopus (Elsevier)

Table 6 shows the SI for each KA. The EU27 was specialised in only KA4 and KA5, while the MIMs were slightly more specialised than the world level in all KAs except KA1, KA4 and KA5. Once again, Table 6 reveals that the same countries tend to rank highly across all KAs, KA9 again being the exception. Latvia was at or near the top in most KAs, and in fact obtained an outstanding SI in KA4, KA5 and KA7. It is worth mentioning that Latvia's publication output was quite low (162 publications for all KAs combined), making it easier to diverge away from the world level. The countries that were highly specialised in KA9 (nuclear safety) were those that produce a significant share of their electricity with nuclear power, such as South Korea, Japan and France. The MIMs' SI has shown stability between 2015 and 2019 across all KAs, while the EU27's SI decreased in every KA (except KA4).

Table 6: Specialisation index for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EU27	**EU27**	**0.85**	**0.85**	**0.90**	**0.69**	**1.49**	**1.36**	**0.75**	**0.79**	**0.74**	**1.00**
MIM	**MIM**	**1.03**	**0.99**	**1.02**	**1.06**	**0.88**	**0.92**	**1.13**	**1.03**	**1.08**	**1.08**
Latvia	LV	1.87	1.09	1.80	1.67	6.16	3.91	1.72	2.47	0.78	0.53
United Arab Emirates	AE	1.77	1.62	2.41	1.61	4.25	2.97	1.02	1.20	2.32	0.82
Rep. of Korea	KR	1.64	1.42	2.07	1.35	1.36	1.33	2.07	1.48	1.68	2.35
India	IN	1.59	1.66	2.48	1.69	0.73	1.35	0.88	1.72	0.72	0.74
Romania	RO	1.54	1.87	1.69	1.31	2.90	2.72	1.64	1.10	0.67	0.95
China	CN	1.44	1.34	1.17	1.78	0.84	0.90	2.02	1.43	1.58	1.12
Cyprus	CY	1.44	1.85	2.00	1.53	3.79	2.00	0.49	0.69	0.93	0.09
Saudi Arabia	SA	1.31	1.79	2.12	0.90	1.42	1.38	0.40	1.00	2.22	0.30
Estonia	EE	1.21	1.00	1.57	1.25	4.53	1.61	0.43	1.06	0.17	0.49
Denmark	DK	1.17	1.73	0.69	1.36	2.02	1.38	0.72	1.14	0.76	0.31
Greece	EL	1.09	1.08	1.80	1.00	1.87	1.23	0.61	0.91	0.65	0.29
Portugal	PT	1.07	1.19	1.31	0.81	2.53	2.30	0.51	1.03	0.83	0.46
Indonesia	ID	1.02	1.43	0.99	0.50	1.16	1.52	0.65	1.47	0.63	0.89
Finland	FI	1.02	0.68	1.44	0.78	1.53	2.37	0.71	0.97	0.56	1.73
Norway	NO	1.01	1.58	0.78	0.61	2.19	1.20	0.50	0.66	3.19	0.27
Slovenia	SI	0.97	0.80	0.98	0.71	2.75	1.89	0.77	0.79	0.52	1.76
Slovakia	SK	0.96	0.65	1.13	0.64	3.46	2.77	1.09	0.69	0.44	0.90
Lithuania	LT	0.95	0.82	1.14	0.44	1.73	1.39	0.49	1.41	0.50	1.62
Sweden	SE	0.94	0.89	0.87	0.60	1.82	1.83	0.91	1.03	0.97	1.20
Italy	IT	0.93	0.72	1.09	0.80	2.02	1.54	0.86	0.84	0.77	0.97
Luxembourg	LU	0.92	1.00	2.21	0.52	1.72	1.29	0.86	0.29	0.00	0.17
Bulgaria	BG	0.89	0.75	1.35	0.76	0.91	1.01	1.27	0.74	0.35	0.63
Mexico	MX	0.87	1.23	0.69	0.51	0.59	1.39	0.25	1.57	0.90	0.44
Ireland	IE	0.85	0.99	1.51	0.71	1.36	1.41	0.33	0.62	0.55	0.06
Germany	DE	0.84	0.92	0.78	0.65	0.69	1.53	1.07	0.72	0.84	1.13
Japan	JP	0.83	0.82	0.73	0.61	0.60	0.44	0.84	0.87	0.80	2.17
Czech Republic	CZ	0.82	0.42	0.79	0.55	3.00	0.89	0.59	0.90	0.53	1.92
Croatia	HR	0.81	0.94	0.84	1.01	1.03	1.43	0.59	0.65	0.31	0.35
Spain	ES	0.79	1.00	0.79	0.64	1.39	1.04	0.49	0.76	0.84	0.56
Austria	AT	0.78	0.51	1.06	0.65	1.64	1.98	0.68	0.77	0.51	0.49
Brazil	BR	0.78	0.69	0.69	0.79	0.49	1.29	0.26	1.47	0.69	0.22
Poland	PL	0.77	0.65	0.57	0.72	1.78	1.38	0.63	1.01	0.74	0.54
Canada	CA	0.75	0.60	0.79	0.74	1.07	0.62	0.83	0.74	1.16	0.58
Malta	MT	0.74	1.81	0.25	1.25	0.43	0.00	0.13	0.00	0.00	0.92
France	FR	0.72	0.66	0.72	0.53	0.86	0.66	0.58	0.64	0.57	2.06
Australia	AU	0.71	0.76	0.72	0.69	1.18	0.71	0.58	0.68	1.22	0.15
United Kingdom	UK	0.71	0.77	0.72	0.55	1.07	0.86	0.63	0.59	0.94	0.90
Belgium	BE	0.71	0.72	0.91	0.55	1.29	0.71	0.66	0.42	0.92	0.89
Hungary	HU	0.64	0.46	0.66	0.31	1.81	1.30	0.60	0.56	0.89	1.26
United States	US	0.56	0.49	0.53	0.52	0.53	0.46	0.62	0.52	0.74	1.05
Chile	CL	0.56	0.80	0.42	0.74	0.95	0.81	0.24	0.50	0.24	0.09
Netherlands	NL	0.56	0.66	0.55	0.43	0.91	0.69	0.40	0.55	0.78	0.47

Source: Statistics calculated using data from Scopus (Elsevier)

3.3 Share of international co-publications

Box 3-3: Conclusion regarding the share of international co-publications for inclusion in the CEII

> Although it was shown that smaller countries tend to proportionately collaborate relatively more frequently internationally, following these analyses we conclude that the SIP is a pertinent indicator to measure the level of collaboration in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII. The SIP, when normalised against the world weighted average, enables comparison between countries across KAs. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

The share of international publications (SIP) is calculated based on the number of international co-publications and the total number of publications in clean energy research. The SIP is shown in **Error! Reference source not found.** for 2019 and all KAs combined and is normalised according to the world weighted average (the "score" column). Larger economies with an extensive network of universities and research centres are for the most part self-sufficient in terms of research needs and, as a result, collaborate relatively less frequently than smaller economies. This dynamic is also at play among the EU27 members and MIMs in clean energy research, with a negative but weak correlation between GDP share and SIP. The United States, China and India, the world's largest economies, figure among the countries that collaborated internationally proportionally the least. The fact that they are also MIMs explains in large part why the MIMs' SIP was below the world average, even though Saudi Arabia, a MIM, was the country that collaborated internationally relatively the most, along with the United Arab Emirates and Australia, which also figure in Cluster 1. The EU27's collective SIP was above the world average. All EU27 members did obtain an SIP above the world average, except for a few of the smaller members such as Romania, Latvia, Bulgaria, Poland, Slovakia and Malta. The EU27's and the MIMs' SIPs relative to the world weighted average were stable between 2015 and 2019.

Both the number of international co-publications and the total number of publications are computed using a full count at country level but fractioned between KAs. The outlier and break in time series tests were first applied to the number of international co-publications, which flagged a single potential outlier for all KAs combined: Canada in 2019. Following an investigation, the data point was deemed to fall in line with the overall trend and was kept as is. The outlier test did identify a few potential outliers in the different KAs. However, the number of international co-publications at the country and KA levels is often low, and thus prone to yearly fluctuations, which was the basis to dismiss many of these flags. In other instances, the potential outliers were deemed to be in line with the overall trends. The total number of publications in clean energy research was also subjected to the test. It raised many of the same flags that were identified in Section 3.1 for the number of publications in fractional count. In the end, all data points were deemed appropriate. The break in time series was also run on both series and found no unusual behaviour. Finally, the tests were run on the SIP itself and normalised by the world weighted average. The outlier test identified one potential outlier for all the KAs combined: Austria's SIP relative to the world weighted average in 2015, which was not excessively out of line from the overall trend. It also identified a few potential outliers for some countries in some KAs, but only one for a country with a substantial number of publications: Germany's SIP in KA4 in 2015, which was not significant at the 99 % confidence level. Following a visual inspection, it was decided to leave these data points as they were. The break in time series test identified some potential problematic breaks in the trends, none of which were deemed worthy of a correction.

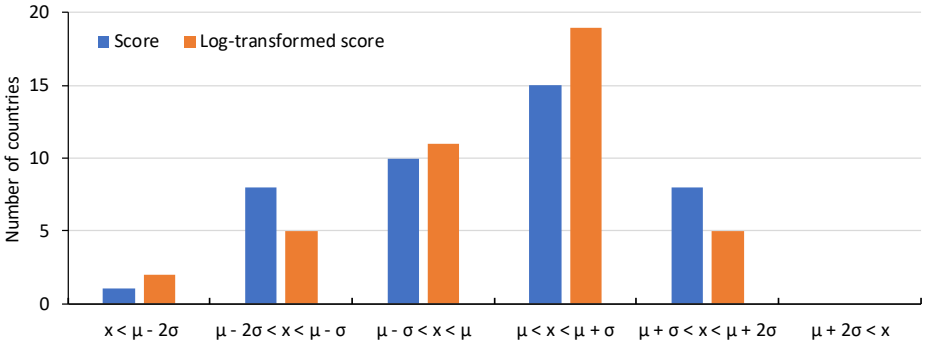
Table 7: Share of international co-publications for all KAs combined

Country	Code	Share in GDP (2019)	SIP (2019)	Score (2019)	Score 0-1 (2019)	CAGR		
						2015 to 2019	Gap to EU27	Gap to MIM
World	**World**	**100%**	**40.4%**	**1.00**	**0.36**	**0.0%**	**-0.4%**	**0.0%**
EU27	**EU27**	**15.3%**	**53.8%**	**1.33**	**0.55**	**0.4%**	**0.0%**	**0.4%**
MIM	**MIM**	**69.3%**	**37.5%**	**0.93**	**0.31**	**0.0%**	**-0.4%**	**0.0%**
Cluster 1		**3.1%**						
Saudi Arabia	SA	1.2%	77.0%	1.91	0.88	-2.5%	-2.9%	-2.5%
Luxembourg	LU	0.1%	71.4%	1.77	0.80	-4.2%	-4.6%	-4.2%
United Arab Emirates	AE	0.5%	69.1%	1.71	0.77	-1.3%	-1.7%	-1.3%
Australia	AU	1.0%	68.3%	1.69	0.76	3.2%	2.8%	3.2%
Denmark	DK	0.3%	68.2%	1.69	0.75	2.2%	1.8%	2.2%
Estonia	EE	0.0%	66.5%	1.65	0.73	2.1%	1.7%	2.1%
Cluster 2		**14.9%**						
Netherlands	NL	0.8%	66.3%	1.64	0.73	0.1%	-0.3%	0.1%
Belgium	BE	0.5%	66.3%	1.64	0.73	0.1%	-0.3%	0.1%
Chile	CL	0.4%	66.2%	1.64	0.73	-0.9%	-1.3%	-0.9%
United Kingdom	UK	2.4%	65.3%	1.62	0.71	2.5%	2.1%	2.5%
Finland	FI	0.2%	65.3%	1.62	0.71	2.7%	2.3%	2.7%
France	FR	2.4%	64.3%	1.59	0.70	1.2%	0.8%	1.2%
Cyprus	CY	0.0%	63.8%	1.58	0.69	2.4%	2.0%	2.4%
Sweden	SE	0.4%	63.1%	1.56	0.68	-1.0%	-1.4%	-1.0%
Lithuania	LT	0.1%	60.5%	1.50	0.64	7.6%	7.2%	7.6%
Ireland	IE	0.3%	59.5%	1.47	0.63	-1.5%	-1.9%	-1.5%
Norway	NO	0.3%	59.4%	1.47	0.63	0.6%	0.2%	0.6%
Austria	AT	0.4%	55.6%	1.38	0.57	-0.8%	-1.2%	-0.8%
Portugal	PT	0.3%	55.4%	1.37	0.57	1.4%	1.0%	1.4%
Canada	CA	1.4%	55.2%	1.37	0.57	1.4%	1.0%	1.4%
Spain	ES	1.5%	54.3%	1.34	0.56	-0.8%	-1.2%	-0.8%
Hungary	HU	0.2%	54.1%	1.34	0.55	3.1%	2.7%	3.1%
Germany	DE	3.4%	50.9%	1.26	0.51	0.9%	0.5%	0.9%
Cluster 3		**24.6%**						
Slovenia	SI	0.1%	49.7%	1.23	0.49	-3.4%	-3.8%	-3.4%
Greece	EL	0.2%	47.9%	1.18	0.46	-3.2%	-3.6%	-3.2%
United States	US	15.8%	46.9%	1.16	0.45	3.7%	3.3%	3.7%
Croatia	HR	0.1%	46.8%	1.16	0.45	9.5%	9.1%	9.5%
Italy	IT	2.0%	46.7%	1.16	0.45	0.8%	0.4%	0.8%
Czech Republic	CZ	0.3%	46.4%	1.15	0.44	2.2%	1.8%	2.2%
Mexico	MX	1.9%	39.8%	0.98	0.35	-1.1%	-1.5%	-1.1%
Malta	MT	0.0%	37.9%	0.94	0.32	-10.9%	-11.3%	-10.9%
Japan	JP	4.0%	36.9%	0.91	0.31	5.1%	4.7%	5.1%
Slovakia	SK	0.1%	36.3%	0.90	0.30	7.7%	7.3%	7.7%
Cluster 4		**32.9%**						
Poland	PL	1.0%	35.0%	0.87	0.28	5.3%	4.9%	5.3%
Bulgaria	BG	0.1%	34.7%	0.86	0.27	-12.6%	-13.0%	-12.6%
Brazil	BR	2.4%	32.1%	0.79	0.24	0.7%	0.3%	0.7%
Latvia	LV	0.0%	31.9%	0.79	0.23	5.8%	5.4%	5.8%
Rep. of Korea	KR	1.7%	30.0%	0.74	0.21	1.4%	1.0%	1.4%
Romania	RO	0.4%	27.4%	0.68	0.17	-0.6%	-1.0%	-0.6%
China	CN	17.7%	23.6%	0.58	0.11	2.3%	1.9%	2.3%
Indonesia	ID	2.5%	20.6%	0.51	0.07	-15.2%	-15.6%	-15.2%
India	IN	7.1%	17.2%	0.43	0.02	-1.3%	-1.7%	-1.3%

Source: Statistics calculated using data from Scopus (Elsevier) and the World Bank

The SIP does not behave like the indicators discussed so far in the sense that it is slightly skewed toward higher ranking countries (**Error! Reference source not found.**). Applying the logarithm to the SIP skews the distribution even more. For this reason, it was not transformed, which is why the log-transformed score is not included in **Error! Reference source not found.**. The standardisation on a 0-to-1 scale was done directly on the score.

Figure 4: Distribution of countries according to the mean and standard deviation of the share of international co-publications normalised by the world weighted average and the log-transformed equivalent (2019)



Source: Statistics calculated using data from Scopus (Elsevier)

Error! Reference source not found. shows the SIP relative to the world weighted average for each KA. The EU27's SIP was above the world weighted average in every KA, while the opposite was true for the MIMs (except for KA4, for which they were on par). Also, the countries that collaborated internationally on many of their publications tended to do so across all KAs, even in KA9. This suggests that the size of an economy is more predictive of its propensity to collaborate internationally then the research topic at hand, in clean energy research at least. Trends were fairly stable for both EU27 members and MIMs in all KAs. The only change worth mentioning is the EU27's 3.8 % increase between 2015 and 2019 in KA6.

Table 8: Share of international co-publications normalised by the world weighted average for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EU27	**EU27**	**1.33**	**1.33**	**1.23**	**1.47**	**1.04**	**1.10**	**1.33**	**1.30**	**1.28**	**1.61**
MIM	**MIM**	**0.93**	**0.93**	**0.93**	**0.90**	**1.00**	**0.96**	**0.95**	**0.93**	**0.95**	**0.94**
Saudi Arabia	SA	1.91	1.96	1.74	2.20	1.69	1.63	2.34	1.90	1.55	1.86
Luxembourg	LU	1.77	1.83	1.32	2.27	1.68	1.49	2.66	1.84	N/C	2.24
United Arab Emirates	AE	1.71	1.63	1.77	2.17	1.24	1.53	1.52	1.78	1.37	1.50
Australia	AU	1.69	1.55	1.60	1.84	1.47	1.70	2.22	1.66	1.57	1.82
Denmark	DK	1.69	1.53	1.67	2.12	1.38	1.35	1.96	1.53	1.76	1.92
Estonia	EE	1.65	1.50	1.50	1.84	1.51	2.13	2.44	1.68	1.72	1.65
Netherlands	NL	1.64	1.60	1.56	1.76	1.52	1.77	1.61	1.60	1.45	1.78
Belgium	BE	1.64	1.59	1.36	1.75	1.28	1.83	1.81	1.79	1.42	1.90
Chile	CL	1.64	1.49	1.45	1.74	1.68	1.58	2.37	1.69	1.67	2.24
United Kingdom	UK	1.62	1.61	1.63	1.88	1.39	1.38	1.69	1.56	1.44	1.45
Finland	FI	1.62	1.60	1.63	1.87	1.55	1.24	1.81	1.36	1.38	1.74
France	FR	1.59	1.59	1.55	1.80	1.42	1.41	1.66	1.55	1.60	1.42
Cyprus	CY	1.58	1.59	1.67	1.72	1.63	1.55	0.50	1.26	1.43	2.24
Sweden	SE	1.56	1.65	1.61	1.87	1.08	1.36	1.62	1.41	1.44	1.49
Lithuania	LT	1.50	1.83	1.46	1.30	1.63	1.31	1.21	1.40	0.78	1.22
Ireland	IE	1.47	1.38	1.35	1.61	1.38	1.05	1.74	1.63	1.78	2.24
Norway	NO	1.47	1.30	1.57	1.85	1.11	1.61	1.75	1.63	1.02	2.00
Austria	AT	1.38	1.53	1.11	1.51	1.13	0.87	1.44	1.46	1.41	1.87
Portugal	PT	1.37	1.29	1.20	1.75	0.95	1.12	1.94	1.28	1.17	1.99
Canada	CA	1.37	1.33	1.42	1.52	0.99	1.58	1.54	1.33	1.21	1.15
Spain	ES	1.34	1.29	1.25	1.49	1.14	1.30	1.58	1.32	1.09	1.65
Hungary	HU	1.34	1.21	0.91	1.67	1.03	1.29	1.41	1.48	0.93	1.63
Germany	DE	1.26	1.26	1.02	1.27	1.19	0.79	1.17	1.31	1.24	1.73
Slovenia	SI	1.23	1.18	1.39	1.16	0.73	0.98	1.48	1.18	1.20	1.45
Greece	EL	1.18	1.18	1.20	1.20	0.98	0.97	1.19	1.23	1.20	1.64
United States	US	1.16	1.16	1.20	1.28	1.01	1.19	1.25	1.20	1.04	0.85
Croatia	HR	1.16	1.16	1.03	1.14	1.13	1.42	0.97	1.30	0.24	1.75
Italy	IT	1.16	1.23	1.07	1.24	0.90	1.05	1.12	1.05	1.16	1.58
Czech Republic	CZ	1.15	1.04	1.21	1.19	0.51	1.31	1.05	1.19	1.17	1.44
Mexico	MX	0.98	0.85	0.91	1.31	0.99	1.24	1.17	0.95	0.86	0.35
Malta	MT	0.94	0.89	1.97	0.77	1.00	N/C	0.00	N/C	N/C	1.12
Japan	JP	0.91	0.94	0.86	0.99	1.03	1.01	0.87	0.96	0.91	0.71
Slovakia	SK	0.90	0.83	0.94	0.88	0.54	0.71	0.41	1.14	1.17	1.45
Poland	PL	0.87	0.80	0.94	0.98	0.41	0.81	0.71	0.81	0.77	1.57
Bulgaria	BG	0.86	0.83	0.36	0.89	0.63	0.68	0.66	1.25	1.49	1.61
Brazil	BR	0.79	0.69	0.99	0.83	1.04	0.82	0.91	0.65	0.86	0.98
Latvia	LV	0.79	1.14	0.67	0.80	0.46	0.52	0.90	0.71	1.29	1.80
Rep. of Korea	KR	0.74	0.83	0.65	0.79	0.56	0.71	0.71	0.87	0.70	0.59
Romania	RO	0.68	0.57	0.53	0.65	0.51	0.59	0.63	0.87	0.93	1.40
China	CN	0.58	0.56	0.70	0.50	0.75	0.69	0.63	0.62	0.62	0.47
Indonesia	ID	0.51	0.42	0.45	0.74	0.46	0.44	0.50	0.57	0.52	0.36
India	IN	0.43	0.43	0.34	0.39	0.64	0.40	0.55	0.51	0.66	0.41

Source: Statistics calculated using data from Scopus (Elsevier)

3.4 Share of transnational co-publications

Box 3-4: Conclusion regarding the share of transnational co-publications for inclusion in the CEII

> Following these analyses, we conclude that the STP is a pertinent indicator to measure the level of collaboration in clean energy research, one of the scientific aspects of interest to the Commission. However, the fact that it is only calculated for EU27 members, and given that it correlates strongly with the SIP, leads us to conclude that it adds little value and therefore should not be included in the CEII.

The share of transnational co-publications (STP) is calculated based on the number of transnational co-publications and the total number of publications in clean energy research. The STP is computed for EU27 members only and is shown in Table 9, along with the normalised score based on the EU27 weighted average. As was the case for international co-publications, larger members of the EU27 collaborated proportionately less frequently with other members. There is indeed a negative but small correlation between GDP share and STP. The EU's largest economies, Germany, France and Italy, all sit at the lower end of Cluster 3 in Table 9. The STP is naturally smaller than the SIP, given that the STP considers a much smaller pool of countries that the EU27 members collaborated with; however, the SIP and the STP are strongly correlated. Luxembourg and Estonia both figure in Cluster 1 of both the SIP and the STP, while Poland, Romania and Latvia all figure in Cluster 4 of both indicators. Denmark is somewhat of an oddity since it figures in Cluster 1 of the SIP distribution, with 68.2 % of its publications written in collaboration with an international partner (**Error! Reference source not found.**). Of that share, only 28.1 % were other EU27 members, enough for a placing in Cluster 3.

The number of transnational co-publications and the total number of publications are both computed using a full count at country level but fractioned between KAs, the latter of which was already tested in Section 3.3. The outlier test identified a single instance of a potential outlier for an EU27 member with a substantial number of publications: Spain's number of transnational co-publications in KA3 in 2015, but it was once again deemed an appropriate data point. The test run on the STP and normalised by the EU27 weighted average flagged Austria once again (its STP relative to the EU27 weighted average for all KAs combined in 2015), but it did not fall excessively outside the overall trend. It also identified a few other potential outliers for some countries in some KAs, but none for a country with a substantial number of publications. The break in time series test identified some potential problematic breaks in the trends, none of which were deemed worthy of a correction.

Table 9: Share of transnational co-publications for all KAs combined

Country	Code	Share in GDP (2019)	STP (2019)	Score (2019)	Score 0-1 (2019)	CAGR	
						2015 to 2019	Gap to EU27
EU27	**EU27**	**15.3%**	**28.7%**	**1.00**	**0.26**	**0.0%**	**0.0%**
Cluster 1		**1.7%**					
Luxembourg	LU	0.1%	46.7%	1.62	0.58	-9.8%	-9.8%
Belgium	BE	0.5%	44.8%	1.56	0.54	0.6%	0.6%
Estonia	EE	0.0%	41.9%	1.46	0.49	-2.0%	-2.0%
Austria	AT	0.4%	40.4%	1.41	0.47	1.9%	1.9%
Cyprus	CY	0.0%	40.3%	1.40	0.47	1.0%	1.0%
Netherlands	NL	0.8%	40.0%	1.39	0.46	1.1%	1.1%
Cluster 2		**0.4%**					
Finland	FI	0.2%	37.6%	1.31	0.42	2.0%	2.0%
Slovenia	SI	0.1%	37.5%	1.31	0.42	-1.5%	-1.5%
Lithuania	LT	0.1%	37.3%	1.30	0.41	11.9%	11.9%
Malta	MT	0.0%	34.5%	1.20	0.36	-0.7%	-0.7%
Cluster 3		**11.7%**					
Sweden	SE	0.4%	31.3%	1.09	0.31	-1.5%	-1.5%
Hungary	HU	0.2%	31.2%	1.09	0.31	3.6%	3.6%
Portugal	PT	0.3%	30.6%	1.06	0.29	-1.5%	-1.5%
Ireland	IE	0.3%	30.0%	1.05	0.28	-2.2%	-2.2%
Czech Republic	CZ	0.3%	29.7%	1.03	0.28	5.4%	5.4%
Spain	ES	1.5%	29.1%	1.01	0.27	-0.1%	-0.1%
Greece	EL	0.2%	29.0%	1.01	0.27	-2.4%	-2.4%
Croatia	HR	0.1%	28.7%	1.00	0.26	12.1%	12.1%
Denmark	DK	0.3%	28.1%	0.98	0.25	-1.1%	-1.1%
Slovakia	SK	0.1%	27.4%	0.96	0.24	9.3%	9.3%
Italy	IT	2.0%	26.6%	0.93	0.22	1.1%	1.1%
France	FR	2.4%	26.5%	0.92	0.22	-2.1%	-2.1%
Germany	DE	3.4%	24.9%	0.87	0.19	0.3%	0.3%
Bulgaria	BG	0.1%	23.9%	0.83	0.18	-14.2%	-14.2%
Cluster 4		**1.5%**					
Poland	PL	1.0%	19.3%	0.67	0.18	1.7%	1.7%
Romania	RO	0.4%	18.6%	0.65	0.08	0.2%	0.2%
Latvia	LV	0.0%	17.6%	0.61	0.07	-0.3%	-0.3%

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

The distribution of the STP (not shown) is skewed toward below-average countries, but it does not exhibit a clear trend given that it is based on a small sample of 27 countries. For that reason, it was not transformed and the standardisation on a 0-to-1 scale was done directly on the score.

Table 10 shows the STP relative to the EU27 weighted average for each KA. As was the case with the SIP, the countries that collaborated with EU27 members on many of their publications tended to do so across all KAs. Luxembourg, which obtained the highest STP for all KAs combined, obtained the highest STP in KA3, KA5, KA7 and KA9. Estonia also obtained the highest STP in KA1, KA4 and KA6. Although, it should be noted that the publication output of these two countries was low.

Table 10: Share of transnational co-publications normalised by the EU27 weighted average for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
EU27	**EU27**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Luxembourg	LU	1.62	1.38	1.46	2.08	1.56	2.46	2.03	2.72	N/C	1.90
Belgium	BE	1.56	1.64	1.41	1.48	1.39	1.73	1.87	1.60	1.43	1.41
Estonia	EE	1.46	1.65	1.21	1.35	2.20	1.97	2.26	1.43	1.41	1.27
Austria	AT	1.41	1.45	1.31	1.50	1.49	1.19	1.85	1.61	1.22	1.24
Cyprus	CY	1.40	1.40	1.69	1.52	1.91	0.93	0.79	1.34	1.76	1.27
Netherlands	NL	1.39	1.52	1.36	1.40	1.32	1.72	1.60	1.41	1.32	1.18
Finland	FI	1.31	1.15	1.33	1.68	1.55	1.48	0.98	1.16	1.31	1.06
Slovenia	SI	1.31	1.42	0.90	1.35	1.05	1.50	1.77	1.51	1.56	1.10
Lithuania	LT	1.30	1.47	1.70	0.82	1.87	1.71	1.00	1.42	0.00	0.52
Malta	MT	1.20	1.35	3.15	0.65	1.58	N/C	0.00	N/C	N/C	0.95
Sweden	SE	1.09	1.10	1.30	1.25	0.97	1.06	1.37	0.94	0.89	0.94
Hungary	HU	1.09	0.72	0.95	1.02	0.87	1.65	1.17	1.22	0.58	1.06
Portugal	PT	1.06	1.12	0.64	1.47	0.96	0.94	1.67	0.98	0.81	1.58
Ireland	IE	1.05	1.27	1.08	0.92	1.08	1.12	1.34	1.20	0.66	1.69
Czech Republic	CZ	1.03	0.84	1.13	0.78	0.55	1.23	0.77	1.24	0.83	1.01
Spain	ES	1.01	0.99	1.01	1.05	1.05	1.01	1.48	0.93	0.90	1.18
Greece	EL	1.01	0.91	1.35	0.77	1.27	1.11	1.23	1.09	1.19	0.88
Croatia	HR	1.00	0.87	1.20	0.75	1.68	1.75	1.03	1.14	0.39	1.18
Denmark	DK	0.98	0.96	1.36	0.82	1.20	1.35	1.02	1.08	1.23	1.13
Slovakia	SK	0.96	0.91	1.14	1.01	0.69	0.98	0.41	1.11	0.96	1.08
Italy	IT	0.93	0.95	0.87	0.97	0.90	0.92	0.84	0.93	1.02	1.04
France	FR	0.92	0.86	0.76	0.93	0.85	0.78	0.95	0.94	1.23	0.75
Germany	DE	0.87	0.91	0.84	0.80	1.11	0.68	0.70	0.85	0.78	0.95
Bulgaria	BG	0.83	1.08	0.30	0.99	0.33	1.11	0.48	1.18	2.43	0.86
Poland	PL	0.67	0.61	0.89	0.65	0.44	0.75	0.46	0.60	0.66	1.08
Romania	RO	0.65	0.48	0.63	0.71	0.65	0.51	0.55	0.81	1.45	0.90
Latvia	LV	0.61	0.46	0.78	0.26	0.51	0.86	0.80	0.90	0.00	1.14

Source: These statistics were calculated using data from Scopus (Elsevier)

3.5 Weighted eigenvector centrality in the world's co-publication network

Box 3-5: Conclusion regarding the weighted eigenvector centrality in the world's co-publication network for inclusion in the CEII

> Following these analyses, we conclude that the WEC within the world's co-publication network is not a particularly pertinent indicator to measure the level of collaboration in clean energy research, for the simple reason that it is highly correlated with the number of publications. Large economies tend to rank high even if they collaborated relatively less frequently with international partners than smaller economies. We do not recommend including it in the CEII.

The weighted eigenvector centrality (WEC)[5] was calculated twice: once for the EU27 members and MIMs among the world's co-publication network, which is shown in Table 11 for 2019 and all KAs combined, and once for the EU27 members among their own co-publication network, which is discussed in Section 3.6. The WEC is heavily skewed toward lower ranking countries, which are also the smallest economies. There is in fact a very strong correlation between the WEC and the number of publications, which is understandable given that countries who publish many articles also tend to publish many international co-publications, and as a result have diverse collaboration partners. China obtained the highest WEC even though only 23.6 % of its clean energy publications were international co-publications, the 3rd lowest share (**Error! Reference source not found.**). The same dynamic is at play in all the other KAs (Table 12). China and the United States obtained the highest and second highest scores in every one of them, except KA9. The countries that ranked high did so in every KA, while the opposite is true for lower ranking countries.

The WEC does not exhibit a clear trend for many countries and its behaviour is even more sporadic at the KA level. The outlier and break in time series tests were nevertheless applied to the score and did identify a few potentially problematic data points. But ultimately, none of them were corrected after inspecting them.

As was mentioned previously, the WEC is heavily skewed toward lower ranking countries. Applying the logarithm does little to flatten the curve to a more "normal-looking" distribution (not shown). In Table 11, the score was log-transformed nonetheless, which was used to standardise it further on a 0-to-1 scale.

[5] The WEC is a network indicator that measures the level of integration of an actor in a collaboration network. It integrates the number of actors to which a given actor is connected through co-authorship of publications, the intensity of those connections (number of co-publications between two actors), and the importance of the partnering actors to the network structure (connections to hubs are valued more than connections to peripheral actors). Scores for this indicator range from 0 to 1, with 1 representing the most important actor to the network structure (typically a major hub with strong connections to a larger number of actors including other influential players in the network) and 0 representing isolated entities (actors which are disconnected from the network).

Table 11: Weighted eigenvector centrality among the world's co-publication network

Country	Code	Share in GDP (2019)	WEC (2019)	Log score (2019)	Score 0-1 (2019)	CAGR (2015 to 2019)
Cluster 1		**36.9%**				
China	CN	17.7%	0.610	0.48	1.00	3.2%
United States	US	15.8%	0.552	0.44	0.92	-0.5%
United Kingdom	UK	2.4%	0.279	0.25	0.52	0.3%
Australia	AU	1.0%	0.210	0.19	0.40	6.2%
Cluster 2		**23.5%**				
Germany	DE	3.4%	0.185	0.17	0.36	-6.8%
Canada	CA	1.4%	0.145	0.14	0.28	-1.6%
Rep. of Korea	KR	1.7%	0.138	0.13	0.27	-4.4%
Japan	JP	4.0%	0.132	0.12	0.26	-4.7%
France	FR	2.4%	0.123	0.12	0.24	-9.3%
Italy	IT	2.0%	0.109	0.10	0.22	-8.3%
India	IN	7.1%	0.109	0.10	0.22	5.1%
Spain	ES	1.5%	0.088	0.08	0.18	-8.8%
Cluster 3		**15.1%**				
Denmark	DK	0.3%	0.073	0.07	0.15	0.7%
Saudi Arabia	SA	1.2%	0.072	0.07	0.15	0.5%
Sweden	SE	0.4%	0.071	0.07	0.14	-6.8%
Netherlands	NL	0.8%	0.065	0.06	0.13	-8.5%
Brazil	BR	2.4%	0.043	0.04	0.09	-2.1%
Finland	FI	0.2%	0.040	0.04	0.08	-2.9%
Belgium	BE	0.5%	0.038	0.04	0.08	-11.2%
Norway	NO	0.3%	0.037	0.04	0.08	-0.6%
Poland	PL	1.0%	0.032	0.03	0.07	5.8%
Portugal	PT	0.3%	0.032	0.03	0.07	-5.9%
Austria	AT	0.4%	0.026	0.03	0.05	-10.4%
United Arab Emirates	AE	0.5%	0.025	0.02	0.05	3.2%
Greece	EL	0.2%	0.023	0.02	0.05	-7.3%
Ireland	IE	0.3%	0.021	0.02	0.04	-6.7%
Czech Republic	CZ	0.3%	0.021	0.02	0.04	-7.4%
Mexico	MX	1.9%	0.018	0.02	0.04	-3.3%
Indonesia	ID	2.5%	0.015	0.01	0.03	13.4%
Romania	RO	0.4%	0.011	0.01	0.02	-3.8%
Hungary	HU	0.2%	0.010	0.01	0.02	-3.0%
Chile	CL	0.4%	0.010	0.01	0.02	-5.7%
Slovenia	SI	0.1%	0.007	0.01	0.01	-4.5%
Croatia	HR	0.1%	0.005	0.01	0.01	0.3%
Cyprus	CY	0.0%	0.005	0.00	0.01	7.9%
Estonia	EE	0.0%	0.005	0.00	0.01	2.6%
Lithuania	LT	0.1%	0.004	0.00	0.01	-6.2%
Slovakia	SK	0.1%	0.004	0.00	0.01	-3.5%
Bulgaria	BG	0.1%	0.003	0.00	0.01	-13.0%
Luxembourg	LU	0.1%	0.002	0.00	0.00	-11.7%
Latvia	LV	0.0%	0.002	0.00	0.00	-6.8%
Malta	MT	0.0%	0.000	0.00	0.00	-2.5%

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Table 12: Weighted eigenvector centrality among the world's co-publication network for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
China	CN	0.61	0.58	0.62	0.64	0.58	0.56	0.66	0.62	0.62	0.28
United States	US	0.55	0.46	0.57	0.56	0.47	0.52	0.57	0.56	0.55	0.40
United Kingdom	UK	0.28	0.40	0.27	0.27	0.36	0.31	0.26	0.22	0.28	0.32
Australia	AU	0.21	0.16	0.20	0.23	0.28	0.20	0.23	0.22	0.28	0.04
Germany	DE	0.18	0.18	0.11	0.12	0.16	0.19	0.14	0.17	0.17	0.42
Canada	CA	0.14	0.14	0.17	0.14	0.12	0.18	0.17	0.15	0.17	0.07
Rep. of Korea	KR	0.14	0.08	0.15	0.11	0.09	0.12	0.10	0.16	0.11	0.12
Japan	JP	0.13	0.10	0.11	0.10	0.14	0.08	0.09	0.14	0.11	0.19
France	FR	0.12	0.10	0.10	0.07	0.11	0.12	0.08	0.10	0.09	0.37
Italy	IT	0.11	0.10	0.10	0.09	0.16	0.18	0.07	0.08	0.08	0.28
India	IN	0.11	0.06	0.14	0.09	0.12	0.10	0.06	0.13	0.07	0.04
Spain	ES	0.09	0.12	0.07	0.07	0.11	0.12	0.05	0.07	0.07	0.20
Denmark	DK	0.07	0.23	0.05	0.12	0.10	0.08	0.06	0.06	0.05	0.05
Saudi Arabia	SA	0.07	0.06	0.09	0.06	0.06	0.07	0.04	0.07	0.08	0.01
Sweden	SE	0.07	0.07	0.07	0.06	0.09	0.13	0.04	0.06	0.05	0.12
Netherlands	NL	0.06	0.10	0.05	0.04	0.09	0.10	0.03	0.06	0.07	0.12
Brazil	BR	0.04	0.05	0.04	0.04	0.04	0.06	0.02	0.06	0.03	0.03
Finland	FI	0.04	0.06	0.05	0.03	0.05	0.07	0.02	0.03	0.02	0.12
Belgium	BE	0.04	0.04	0.02	0.03	0.04	0.08	0.02	0.02	0.03	0.14
Norway	NO	0.04	0.11	0.03	0.03	0.05	0.09	0.02	0.03	0.05	0.03
Poland	PL	0.03	0.02	0.02	0.03	0.03	0.06	0.02	0.03	0.02	0.09
Portugal	PT	0.03	0.06	0.04	0.02	0.05	0.04	0.02	0.02	0.01	0.10
Austria	AT	0.03	0.02	0.02	0.02	0.03	0.04	0.01	0.03	0.03	0.06
United Arab Emirates	AE	0.02	0.02	0.04	0.03	0.04	0.07	0.01	0.02	0.03	0.00
Greece	EL	0.02	0.04	0.03	0.02	0.03	0.02	0.01	0.02	0.02	0.03
Ireland	IE	0.02	0.04	0.02	0.02	0.04	0.03	0.01	0.02	0.02	0.02
Czech Republic	CZ	0.02	0.02	0.02	0.01	0.02	0.03	0.01	0.02	0.02	0.11
Mexico	MX	0.02	0.02	0.01	0.01	0.01	0.07	0.01	0.03	0.01	0.00
Indonesia	ID	0.01	0.03	0.01	0.01	0.02	0.02	0.01	0.03	0.01	0.00
Romania	RO	0.01	0.02	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.05
Hungary	HU	0.01	0.02	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.06
Chile	CL	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00
Slovenia	SI	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.04
Croatia	HR	0.01	0.01	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.01
Cyprus	CY	0.00	0.02	0.01	0.01	0.02	0.01	0.00	0.00	0.00	0.00
Estonia	EE	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Lithuania	LT	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Slovakia	SK	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02
Bulgaria	BG	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
Luxembourg	LU	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	N/C	0.01
Latvia	LV	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Malta	MT	0.00	0.01	0.00	0.00	0.00	N/C	N/C	N/C	N/C	0.00

Source: These statistics were calculated using data from Scopus (Elsevier)

3.6 Weighted eigenvector centrality in the Member States' co-publication network

Box 3-6: Conclusion regarding the weighted eigenvector centrality in the Member States' co-publication network for inclusion in the CEII

> Aside from the fact that the WEC within the Member States' co-publication network is only calculated for EU27 members, it is not a pertinent indicator to measure the level of collaboration in clean energy research, for the same reason as discussed in Section 3.5. We do not recommend including it in the CEII.

The WEC among the EU27 members is shown in **Error! Not a valid bookmark self-reference.**. The same conclusions can be drawn, regardless of the network: the EU27's largest economies, those that also published the most articles, obtained the highest scores. Once again, the countries that ranked high did so in every KA (Table 14). Germany obtained the highest score in KA1, KA6, KA7, KA8 and KA9, while Italy did the best in KA2, KA3, KA4 and KA5.

The WEC within the EU27 network also exhibits fluctuating trends for many countries. However, the outlier and break in time series tests identified very few potential problematic data points: one for Italy in KA7 in 2018 and one for Greece in KA9 in 2017. Following a visual inspection, these data points were left unchanged.

Finally, the distribution of the WEC within the EU27 network was also heavily skewed toward lower ranking countries (not shown). The logarithm was applied to this indicator, which is shown in **Error! Not a valid bookmark self-reference.** and was used to standardise the indicator on a 0-to-1 scale, also shown in **Error! Not a valid bookmark self-reference.**.

Table 13:Weighted eigenvector centrality among the EU27's co-publication network

Country	Code	Share in GDP (2019)	WEC (2019)	Log score (2019)	Score 0-1 (2019)	CAGR (2015 to 2019)
Cluster 1		**9.3%**				
Germany	DE	3.4%	0.488	0.40	0.96	-0.9%
Italy	IT	2.0%	0.428	0.36	0.86	0.5%
France	FR	2.4%	0.385	0.33	0.79	-3.5%
Spain	ES	1.5%	0.378	0.32	0.77	1.0%
Cluster 2		**2.8%**				
Netherlands	NL	0.8%	0.260	0.23	0.56	1.8%
Belgium	BE	0.5%	0.198	0.18	0.43	-1.4%
Sweden	SE	0.4%	0.191	0.17	0.42	0.0%
Denmark	DK	0.3%	0.167	0.15	0.37	6.0%
Portugal	PT	0.3%	0.153	0.14	0.34	1.4%
Austria	AT	0.4%	0.143	0.13	0.32	3.7%
Finland	FI	0.2%	0.137	0.13	0.31	4.3%
Cluster 3		**3.2%**				
Poland	PL	1.0%	0.123	0.12	0.28	11.9%
Greece	EL	0.2%	0.104	0.10	0.24	2.4%
Czech Republic	CZ	0.3%	0.101	0.10	0.23	6.8%
Ireland	IE	0.3%	0.069	0.07	0.16	4.3%
Romania	RO	0.4%	0.068	0.07	0.15	8.2%
Slovenia	SI	0.1%	0.049	0.05	0.11	11.6%
Hungary	HU	0.2%	0.048	0.05	0.11	11.1%
Slovakia	SK	0.1%	0.027	0.03	0.06	17.6%
Croatia	HR	0.1%	0.024	0.02	0.05	12.8%
Cyprus	CY	0.0%	0.020	0.02	0.04	22.5%
Estonia	EE	0.0%	0.020	0.02	0.04	10.8%
Lithuania	LT	0.1%	0.018	0.02	0.04	12.3%
Bulgaria	BG	0.1%	0.016	0.02	0.03	-1.1%
Luxembourg	LU	0.1%	0.014	0.01	0.03	-3.9%
Latvia	LV	0.0%	0.012	0.01	0.02	4.4%
Cluster 4		**0.0%**				
Malta	MT	0.0%	0.004	0.00	0.00	17.0%

Source: These statistics were calculated using data from Scopus (Elsevier)

Table 14: Weighted eigenvector centrality among the EU27's co-publication network for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
Germany	DE	0.49	0.47	0.43	0.45	0.40	0.49	0.51	0.48	0.46	0.51
Italy	IT	0.43	0.28	0.48	0.47	0.53	0.50	0.40	0.41	0.43	0.41
France	FR	0.39	0.25	0.32	0.35	0.23	0.25	0.37	0.37	0.42	0.47
Spain	ES	0.38	0.32	0.41	0.42	0.43	0.35	0.40	0.37	0.39	0.31
Netherlands	NL	0.26	0.40	0.23	0.23	0.20	0.28	0.25	0.32	0.36	0.18
Belgium	BE	0.20	0.20	0.20	0.14	0.15	0.18	0.24	0.13	0.19	0.23
Sweden	SE	0.19	0.19	0.22	0.18	0.23	0.21	0.25	0.19	0.14	0.16
Denmark	DK	0.17	0.40	0.17	0.24	0.24	0.16	0.15	0.20	0.16	0.08
Portugal	PT	0.15	0.19	0.11	0.18	0.18	0.14	0.16	0.13	0.11	0.16
Austria	AT	0.14	0.11	0.14	0.15	0.19	0.22	0.15	0.20	0.11	0.10
Finland	FI	0.14	0.13	0.17	0.16	0.10	0.14	0.08	0.13	0.07	0.17
Poland	PL	0.12	0.10	0.12	0.11	0.10	0.16	0.06	0.12	0.09	0.16
Greece	EL	0.10	0.13	0.19	0.12	0.18	0.10	0.09	0.12	0.11	0.05
Czech Republic	CZ	0.10	0.06	0.08	0.03	0.06	0.04	0.05	0.13	0.09	0.17
Ireland	IE	0.07	0.09	0.11	0.08	0.08	0.10	0.03	0.06	0.06	0.04
Romania	RO	0.07	0.06	0.07	0.08	0.08	0.04	0.04	0.06	0.10	0.07
Slovenia	SI	0.05	0.08	0.03	0.04	0.06	0.01	0.07	0.04	0.01	0.06
Hungary	HU	0.05	0.03	0.03	0.02	0.04	0.05	0.02	0.04	0.02	0.09
Slovakia	SK	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.03	0.01	0.04
Croatia	HR	0.02	0.04	0.04	0.02	0.06	0.03	0.02	0.02	0.00	0.02
Cyprus	CY	0.02	0.11	0.04	0.04	0.07	0.01	0.01	0.01	0.01	0.01
Estonia	EE	0.02	0.06	0.03	0.03	0.04	0.03	0.01	0.02	0.01	0.00
Lithuania	LT	0.02	0.04	0.02	0.00	0.03	0.01	0.00	0.02	N/C	0.02
Bulgaria	BG	0.02	0.03	0.00	0.01	0.01	0.01	0.00	0.02	0.01	0.02
Luxembourg	LU	0.01	0.05	0.03	0.01	0.01	0.01	0.02	0.01	N/C	0.01
Latvia	LV	0.01	0.01	0.01	0.00	0.03	0.01	0.01	0.02	N/C	0.01
Malta	MT	0.00	0.03	0.01	0.01	0.01	N/C	N/C	N/C	N/C	0.00

Source: These statistics were calculated using data from Scopus (Elsevier)

3.7 Share of open access publications

Box 3-7: Conclusion regarding the share of open access publications for inclusion in the CEII

> Following these analyses, we conclude that the SOA is a pertinent indicator to measure the level of research accessibility in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII. The SOA can be normalised against the world level, which enables comparison between countries across KAs. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were ultimately not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

The share of open access publication (SOA) is calculated based on the number of publications available in open access and the total number of publications in clean energy research. The SOA is shown in Table 15 for 2017 and all KAs combined, and it is normalised according to the world weighted average (the "score" column). Globally, 29.0 % of clean energy publications published in 2017 were available in open access. The EU's policy on open access appears to be bearing fruit, since the EU27's SOA was much higher, at 40.9 %. In fact, Greece, Romania and Malta were the only EU27 members to have obtained an SOA below the world average. All countries in Cluster 1 are EU27 members, except for the United Kingdom in 1st position, which was an EU member until recently. Many of these countries' SOAs increased substantially between 2013 and 2017, as did the collective EU27, by 3.5 % annually. Due to delays in making some publications available in open access (due to embargoes and other causes such as growing number of open access repositories), the growth of open access is most likely underestimated; open access data for 2013 to 2017 were all extracted on the same date. In contrast, MIMs tended to have published relatively few of their clean energy publications in open access. China and India, two of the world's largest economies, published fewer than 20 % of their clean energy publications in open access. There was indeed a negative but weak correlation between GDP share and SOA.

The number of publications available in open access and the total number of publications are both computed using a full count at country level but fractioned between KAs. However, as was mentioned in Section 2.1.3, the open access status of publications can only be assessed for those indexed in both Scopus and 1findr. The total number of publications is therefore based on this overlap instead of the total used to calculate the share of international and transnational co-publications. The outlier and break in time series tests were run on that total specifically, as well as on the number of publications available in open access, and on the SOA itself and normalised by the world weighted average. The outlier test run on the total number of publications identified three potential outliers: the Netherlands' output in KA1 in 2015, Germany's output in KA6 in 2015 and France's output in KA8 in 2017. The Netherlands' and Germany's potential outliers were in line with the overall trend, but France's potential outlier was more problematic. After investigating the data in Scopus and 1findr meticulously, it was concluded that the data point was indeed accurate. The test run on the other series flagged a few potential outliers, many of which were in 2019, when a drop was often observed. The year 2019 is not represented in Table 15, but it is worth mentioning, given that these flags are explained by the embargo effect discussed in Section 2.1.3, that being the period of time imposed by some publishers before an article can be made available in open access. Therefore, the drops observed in recent years justify the two-year window imposed on this indicator, which is why Table 15 is based on the period 2013 to 2017 instead of the conventional period of 2015 to 2019. Within the period of interest, a few potential outliers with a substantial number of publications were flagged: Austria's SOA for all KAs combined in 2015, Italy's SOA in KA9 in 2017, India's SOA normalised by the world weighted average in KA2 in 2014 and Spain's SOA normalised by the world weighted

average in KA9 in 2015. Following a visual inspection, it was decided to leave these data points as they were. The break in time series test identified some potential problematic breaks in the trends, none of which were deemed worthy of a correction.

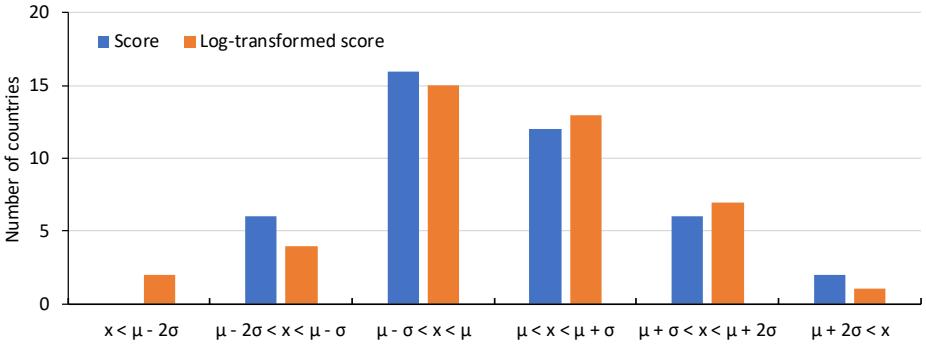
Table 15: Share of open access publications for all KAs combined

Country	Code	Share in GDP (2019)	SOA (2017)	Score (2017)	Log score (2017)	Score 0-1 (2017)	CAGR		
							2013 to 2017	Gap to EU27	Gap to MIM
World	**World**	**100%**	**29.0%**	**1.00**	**0.69**	**0.35**	**0.0%**	**-3.5%**	**0.1%**
EU27	**EU27**	**15.3%**	**40.9%**	**1.41**	**0.88**	**0.58**	**3.5%**	**0.0%**	**3.6%**
MIM	**MIM**	**69.3%**	**27.2%**	**0.94**	**0.66**	**0.32**	**-0.1%**	**-3.6%**	**0.0%**
Cluster 1		**6.6%**							
United Kingdom	UK	2.4%	68.2%	2.35	1.21	0.99	12.3%	8.8%	12.4%
Netherlands	NL	0.8%	62.2%	2.15	1.15	0.91	15.7%	12.2%	15.8%
Luxembourg	LU	0.1%	57.4%	1.98	1.09	0.85	5.6%	2.1%	5.7%
Hungary	HU	0.2%	54.9%	1.90	1.06	0.81	3.5%	0.0%	3.6%
Ireland	IE	0.3%	54.0%	1.86	1.05	0.80	0.8%	-2.7%	0.9%
Denmark	DK	0.3%	53.0%	1.83	1.04	0.78	8.2%	4.7%	8.3%
Croatia	HR	0.1%	50.7%	1.75	1.01	0.75	0.5%	-3.0%	0.6%
Indonesia	ID	2.5%	50.3%	1.74	1.01	0.74	-3.5%	-7.0%	-3.4%
Cluster 2		**7.2%**							
Spain	ES	1.5%	48.3%	1.67	0.98	0.71	5.1%	1.6%	5.2%
Latvia	LV	0.0%	47.0%	1.62	0.96	0.69	-9.1%	-12.6%	-9.0%
Austria	AT	0.4%	46.0%	1.59	0.95	0.67	6.9%	3.4%	7.0%
Chile	CL	0.4%	45.1%	1.56	0.94	0.66	-2.1%	-5.6%	-2.0%
Belgium	BE	0.5%	44.8%	1.55	0.93	0.65	0.4%	-3.1%	0.5%
Norway	NO	0.3%	44.1%	1.52	0.93	0.64	6.0%	2.5%	6.1%
Finland	FI	0.2%	42.3%	1.46	0.90	0.61	6.4%	2.9%	6.5%
Slovenia	SI	0.1%	42.1%	1.45	0.90	0.61	5.2%	1.7%	5.3%
Slovakia	SK	0.1%	39.7%	1.37	0.86	0.56	2.2%	-1.3%	2.3%
Sweden	SE	0.4%	39.4%	1.36	0.86	0.56	-0.2%	-3.7%	-0.1%
Poland	PL	1.0%	39.0%	1.35	0.85	0.55	1.5%	-2.0%	1.6%
France	FR	2.4%	38.8%	1.34	0.85	0.55	2.5%	-1.0%	2.6%
Estonia	EE	0.0%	37.6%	1.30	0.83	0.53	-1.5%	-5.0%	-1.4%
Cluster 3		**33.3%**							
Germany	DE	3.4%	36.4%	1.26	0.81	0.50	3.6%	0.1%	3.7%
Czech Republic	CZ	0.3%	36.1%	1.25	0.81	0.50	-5.3%	-8.8%	-5.2%
Cyprus	CY	0.0%	35.2%	1.21	0.79	0.48	24.8%	21.3%	24.9%
Portugal	PT	0.3%	34.8%	1.20	0.79	0.47	-5.7%	-9.2%	-5.6%
Italy	IT	2.0%	34.3%	1.18	0.78	0.46	4.3%	0.8%	4.4%
Brazil	BR	2.4%	33.1%	1.14	0.76	0.44	-4.3%	-7.8%	-4.2%
Bulgaria	BG	0.1%	33.0%	1.14	0.76	0.44	-0.1%	-3.6%	0.0%
Lithuania	LT	0.1%	32.3%	1.11	0.75	0.42	-7.0%	-10.5%	-6.9%
Japan	JP	4.0%	30.6%	1.06	0.72	0.39	-1.3%	-4.8%	-1.2%
United States	US	15.8%	30.4%	1.05	0.72	0.38	-3.5%	-7.0%	-3.4%
Australia	AU	1.0%	30.0%	1.04	0.71	0.38	-2.6%	-6.1%	-2.5%
Greece	EL	0.2%	28.4%	0.98	0.68	0.34	-0.5%	-4.0%	-0.4%
Saudi Arabia	SA	1.2%	27.6%	0.95	0.67	0.32	-1.0%	-4.5%	-0.9%
Mexico	MX	1.9%	27.5%	0.95	0.67	0.32	-5.0%	-8.5%	-4.9%
Romania	RO	0.4%	26.7%	0.92	0.65	0.31	-3.9%	-7.4%	-3.8%
Cluster 4		**28.5%**							
Malta	MT	0.0%	25.0%	0.86	0.62	0.27	-9.3%	-12.8%	-9.2%
Canada	CA	1.4%	24.4%	0.84	0.61	0.25	-2.9%	-6.4%	-2.8%
United Arab Emirates	AE	0.5%	24.1%	0.83	0.61	0.25	10.2%	6.7%	10.3%
Rep. of Korea	KR	1.7%	20.5%	0.71	0.54	0.16	-2.0%	-5.5%	-1.9%
China	CN	17.7%	16.6%	0.57	0.45	0.06	-0.5%	-4.0%	-0.4%
India	IN	7.1%	15.9%	0.55	0.44	0.04	-3.7%	-7.2%	-3.6%

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

The SOA is slightly skewed toward below-average countries (Figure 5). Applying the logarithm to the SOA corrects the distribution to a more "normal-looking" curve. The log-transformed score is also shown in Table 15 as well as the standardised scores on a 0-to-1 scale, which was calculated based on the log-transformed score.

Figure 5: Distribution of countries according to the mean and standard deviation of the share of open access publications normalised by world weighted average and the log-transformed equivalent (2017)



Source: These statistics were calculated using data from Scopus (Elsevier)

Table 16 shows the SOA relative to the world weighted average for each KA. The EU27's SOA was above the world weighted average in every KA, while the MIMs' share was slightly below the world weighted average in every KA. As was the case for all other indicators, the countries that published a relatively large share of their clean energy publications in open access did so across all KAs. There are a few exceptions: Bulgaria obtained the highest SOA in KA4 (tied with Luxembourg), Lithuania obtained the 2nd highest SOA in KA6 (very close to the Netherlands' top score) and Malta obtained the highest SOA in KA7, while ranking much lower in the other KAs. However, these countries published few articles, making the score more prone to fluctuations. The EU27's SOA increased in every KA between 2013 and 2019, particularly in KA4 (annual increase of 4.7 %), while the MIMs' performance change little, by less than 1 % in every KA.

Table 16: Share of open access publications normalised by the world weighted average for each KA (2017) with corresponding trendline (2013-2017)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EU27	**EU27**	**1.41**	**1.36**	**1.38**	**1.62**	**1.14**	**1.25**	**1.45**	**1.33**	**1.35**	**1.43**
MIM	**MIM**	**0.94**	**0.95**	**0.93**	**0.91**	**0.95**	**0.93**	**0.94**	**0.96**	**0.97**	**0.96**
United Kingdom	UK	2.35	2.38	2.25	2.59	2.08	2.28	2.47	2.30	2.53	1.97
Netherlands	NL	2.15	2.09	1.84	2.75	1.91	1.49	2.80	2.14	2.08	1.93
Luxembourg	LU	1.98	2.02	2.13	1.70	2.80	0.00	1.44	1.78	1.94	2.89
Hungary	HU	1.90	1.82	1.91	1.60	2.07	0.95	2.59	1.69	2.90	1.63
Ireland	IE	1.86	1.86	1.87	2.10	1.61	0.85	2.46	1.76	1.44	2.55
Denmark	DK	1.83	1.73	1.58	2.29	1.22	1.74	2.31	1.78	1.86	1.99
Croatia	HR	1.75	1.66	1.59	1.94	1.98	1.36	1.55	1.58	1.70	2.37
Indonesia	ID	1.74	1.63	1.35	1.96	1.68	1.35	2.59	1.80	1.99	1.26
Spain	ES	1.67	1.72	1.97	2.00	1.45	1.60	1.80	1.44	1.42	1.26
Latvia	LV	1.62	2.84	1.34	2.06	0.84	2.04	0.00	1.15	0.72	2.07
Austria	AT	1.59	1.38	1.66	1.87	1.25	1.15	1.09	1.42	2.10	1.85
Chile	CL	1.56	1.46	1.38	2.01	1.01	1.39	1.50	1.56	1.71	2.10
Belgium	BE	1.55	1.46	1.85	2.12	1.44	1.30	1.28	1.30	0.99	1.69
Norway	NO	1.52	1.65	1.76	2.06	1.15	0.87	1.25	1.64	1.04	1.51
Finland	FI	1.46	1.48	1.49	1.66	0.99	1.13	1.57	1.21	1.47	1.49
Slovenia	SI	1.45	1.50	1.70	1.56	1.24	0.96	1.29	1.40	0.98	1.32
Slovakia	SK	1.37	1.59	1.08	1.43	1.44	0.00	1.36	1.07	1.03	1.47
Sweden	SE	1.36	1.25	1.31	1.42	1.09	1.18	1.25	1.44	1.40	1.46
Poland	PL	1.35	1.45	1.22	1.38	0.87	1.41	1.76	1.39	1.15	1.30
France	FR	1.34	1.10	1.42	1.55	1.04	1.28	1.35	1.30	1.49	1.29
Estonia	EE	1.30	1.28	0.98	1.21	1.32	2.27	1.44	1.11	1.20	1.61
Germany	DE	1.26	1.13	0.92	1.40	0.87	1.08	1.11	1.22	1.37	1.59
Czech Republic	CZ	1.25	1.21	1.62	1.65	0.97	2.08	1.27	0.90	1.37	1.09
Cyprus	CY	1.21	1.26	1.38	1.19	0.82	1.14	0.00	1.15	0.00	2.89
Portugal	PT	1.20	1.14	1.06	1.48	0.87	1.53	1.17	0.98	0.59	1.58
Italy	IT	1.18	1.18	1.21	1.15	0.99	1.05	1.43	1.14	0.94	1.25
Brazil	BR	1.14	1.22	1.17	0.94	1.05	1.22	0.87	1.24	0.79	0.62
Bulgaria	BG	1.14	1.60	1.33	1.63	2.80	0.00	1.99	0.89	1.31	0.34
Lithuania	LT	1.11	1.53	0.64	1.98	0.31	1.95	2.79	0.81	0.00	1.11
Japan	JP	1.06	1.08	1.13	0.88	1.12	1.13	1.38	0.95	0.90	1.01
United States	US	1.05	1.08	1.00	1.07	0.76	0.93	1.05	1.19	1.03	0.88
Australia	AU	1.04	1.06	1.11	1.04	0.86	0.79	1.18	1.15	0.86	1.27
Greece	EL	0.98	0.97	0.94	0.92	0.86	1.19	1.22	0.89	0.73	1.47
Saudi Arabia	SA	0.95	1.05	1.01	0.89	0.71	0.67	0.85	0.90	0.95	0.66
Mexico	MX	0.95	0.89	1.05	1.10	1.15	0.82	1.24	0.91	1.24	0.44
Romania	RO	0.92	0.71	0.84	0.83	0.65	1.31	0.78	0.83	0.12	1.39
Malta	MT	0.86	0.97	N/C	1.58	0.00	0.85	0.00	3.40	N/C	N/C
Canada	CA	0.84	1.03	0.86	0.76	0.66	0.88	0.76	0.82	0.86	0.90
United Arab Emirates	AE	0.83	0.83	1.04	0.83	1.31	1.31	1.08	0.39	0.83	0.00
Rep. of Korea	KR	0.71	0.64	0.82	0.81	1.14	0.75	0.65	0.56	0.60	0.70
China	CN	0.57	0.59	0.67	0.56	0.59	0.46	0.71	0.58	0.59	0.38
India	IN	0.55	0.50	0.53	0.65	0.50	0.47	0.66	0.61	0.53	0.30

Source: These statistics were calculated using data from Scopus (Elsevier)

3.8 Number and share of public/private co-publications

Box 3-8: Conclusion regarding the share of public/private co-publications for inclusion in the CEII

> Following these analyses, we conclude that the SPP is a pertinent indicator to measure the level of collaboration fuelling knowledge transfer from the public sector (particularly academia) to the private sector in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII. The SPP can be normalised against the world level, which enables comparison between countries across KAs. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were ultimately not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

The share of public/private co-publications (SPP) is calculated based on the number of public/private co-publications and the total number of publications in clean energy research. The SPP is shown in Table 17 for 2019 and all KAs combined and is normalised according to the world weighted average (the "score" column). The public and private sectors have different priorities when it comes to scientific research and publishing. It is a top priority for the public sector (particularly higher education), whereas R&D performed by the private sector is generally more focused on higher technology readiness levels targeting economic profitability. That being said, the private sector does participate in scientific research and publishing, particularly if it foresees a potential economic benefit to that research. Additionally, governments are emphasising the need for public/private partnerships to promote knowledge transfer towards innovation (and ultimately socioeconomic returns), as R&D is one of the core mechanisms they rely on to improve standards of living and their economic competitiveness. Data in Table 17 suggest that 10.2 % of clean energy publications were borne out of collaboration between the public and private sectors in 2019 globally. Given that such a small share of publications are public/private partnerships, to avoid excessive fluctuations and scores that are not necessarily representative of typical performance, a minimum of 30 publications was imposed to each combination of country, KA and year. For all KAs combined in 2019, this only affected Malta, but for individual KA, a few more countries were affected (Table 18 further below). The EU27 was successful at implicating the public and private sectors in clean energy research, with 14.3 % of its publications borne out of such a collaboration. The MIMs' share of 11.1 % was also higher than the world weighted average. All countries in Cluster 1 are EU27 members, and all but one country in Cluster 4 are MIMs. All countries in Cluster 1 also saw their score increase between 2015 and 2019, while the opposite was true for countries in Cluster 4, meaning that there was a growing divide between higher and lower ranking countries.

Table 17: Share of public/private co-publications for all KAs combined

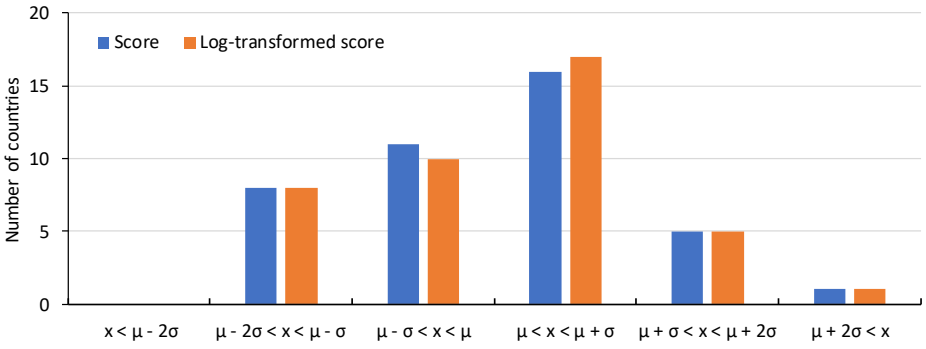
Country	Code	Share in GDP (2019)	SPP (2019)	Score (2019)	Score 0-1 (2019)	CAGR		
						2015 to 2019	Gap to EU27	Gap to MIM
World	**World**	**100%**	**10.2%**	**1.00**	**0.33**	**0.0%**	**-1.3%**	**-0.6%**
EU27	**EU27**	**15.3%**	**14.3%**	**1.41**	**0.50**	**1.3%**	**0.0%**	**0.7%**
MIM	**MIM**	**69.3%**	**11.1%**	**1.09**	**0.37**	**0.6%**	**-0.7%**	**0.0%**
Cluster 1		**4.9%**						
Austria	AT	0.4%	26.0%	2.55	0.96	5.7%	4.4%	5.1%
Netherlands	NL	0.8%	19.9%	1.95	0.72	1.6%	0.3%	1.0%
Croatia	HR	0.1%	19.6%	1.92	0.71	4.3%	3.0%	3.7%
Slovenia	SI	0.1%	18.8%	1.84	0.67	6.5%	5.2%	5.9%
Germany	DE	3.4%	17.8%	1.74	0.63	2.2%	0.9%	1.6%
Finland	FI	0.2%	17.6%	1.72	0.62	1.3%	0.0%	0.7%
Cluster 2		**46.6%**						
Sweden	SE	0.4%	16.8%	1.65	0.60	0.2%	-1.1%	-0.4%
Luxembourg	LU	0.1%	16.2%	1.59	0.57	0.1%	-1.2%	-0.5%
Ireland	IE	0.3%	16.0%	1.56	0.56	-3.8%	-5.1%	-4.4%
Japan	JP	4.0%	15.9%	1.56	0.56	1.3%	0.0%	0.7%
Denmark	DK	0.3%	15.5%	1.52	0.54	-1.1%	-2.4%	-1.7%
Norway	NO	0.3%	15.2%	1.49	0.53	-5.2%	-6.5%	-5.8%
Belgium	BE	0.5%	15.2%	1.49	0.53	1.9%	0.6%	1.3%
France	FR	2.4%	14.9%	1.46	0.52	0.3%	-1.0%	-0.3%
Greece	EL	0.2%	14.6%	1.43	0.51	8.8%	7.5%	8.2%
Cyprus	CY	0.0%	14.5%	1.42	0.50	12.8%	11.5%	12.2%
Italy	IT	2.0%	13.7%	1.34	0.47	0.9%	-0.4%	0.3%
Hungary	HU	0.2%	13.1%	1.28	0.45	6.5%	5.2%	5.9%
Estonia	EE	0.0%	13.2%	1.30	0.45	29.0%	27.7%	28.4%
United Kingdom	UK	2.4%	12.9%	1.27	0.44	-1.6%	-2.9%	-2.2%
United States	US	15.8%	12.0%	1.17	0.40	0.4%	-0.9%	-0.2%
China	CN	17.7%	11.7%	1.15	0.39	7.9%	6.6%	7.3%
Cluster 3		**7.9%**						
Canada	CA	1.4%	11.5%	1.13	0.38	-0.2%	-1.5%	-0.8%
Spain	ES	1.5%	10.6%	1.04	0.35	2.3%	1.0%	1.7%
Rep. of Korea	KR	1.7%	9.7%	0.95	0.31	1.3%	0.0%	0.7%
Czech Republic	CZ	0.3%	8.6%	0.84	0.27	-0.3%	-1.6%	-0.9%
Lithuania	LT	0.1%	8.2%	0.80	0.25	9.5%	8.2%	8.9%
Slovakia	SK	0.1%	8.1%	0.80	0.25	7.6%	6.3%	7.0%
Portugal	PT	0.3%	7.6%	0.75	0.23	4.9%	3.6%	4.3%
Romania	RO	0.4%	7.5%	0.74	0.23	-0.7%	-2.0%	-1.3%
Poland	PL	1.0%	7.2%	0.70	0.21	11.4%	10.1%	10.8%
Bulgaria	BG	0.1%	6.3%	0.62	0.18	-13.4%	-14.7%	-14.0%
Australia	AU	1.0%	6.2%	0.61	0.17	-1.9%	-3.2%	-2.5%
Cluster 4		**16.0%**						
Brazil	BR	2.4%	5.8%	0.57	0.16	-1.2%	-2.5%	-1.8%
United Arab Emirates	AE	0.5%	5.0%	0.49	0.13	-0.9%	-2.2%	-1.5%
Chile	CL	0.4%	5.0%	0.49	0.13	-5.4%	-6.7%	-6.0%
Latvia	LV	0.0%	4.4%	0.43	0.10	-10.9%	-12.2%	-11.5%
Saudi Arabia	SA	1.2%	4.4%	0.43	0.10	-6.1%	-7.4%	-6.7%
Mexico	MX	1.9%	3.2%	0.31	0.05	-1.0%	-2.3%	-1.6%
India	IN	7.1%	2.6%	0.26	0.03	-7.8%	-9.1%	-8.4%
Indonesia	ID	2.5%	2.5%	0.24	0.02	-2.3%	-3.6%	-2.9%
Not calculated		**0.0%**						
Malta	MT	0.0%	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

The number of public/private co-publications and the total number of publications are both computed using a full count at country level but fractioned between KAs. The total number of publications has already been tested for other indicators. The outlier and break in time series tests were therefore applied on the number of public/private co-publications, and on the SPP itself and normalised by the world weighted average. The outlier test run on the number of public/private co-publications identified several potential outliers: Sweden for all KAs combined in 2019, the United Kingdom in KA3 in 2017, Italy in KA3 in 2015, China in KA7 in 2019 and Japan in KA9 in 2017, all of which were not significantly out of line from the overall trend. The test run on the SPP identified no potential outlier for all KAs combined, but it did identify a few for some countries with a substantial number of publications in some KAs: South Korea's SPP in KA1 in 2019, the United States' SPP in KA6 in 2017, the United States' SPP in KA7 in 2019 and India's SPP in KA8 in 2015, none of which were significant at the 99 % confidence level. The test run on the SPP normalised by the world weighted average also identified a few potentially problematic data points. Following a visual inspection, it was decided to leave all these data points as they were. Finally, the break in time series test identified some potential problematic breaks in the trends, none of which were deemed worthy of a correction. Again, it is worth mentioning that public/private collaborations are rare occurrences. As a result, the number of public/private co-publications is often low and can fluctuate significantly from one year to the next, even for countries with a substantial number of publications.

The distribution of the SPP exhibits a "normal-looking" curve but it is slightly skewed toward above-average countries (Figure 6). Applying the logarithm to the SPP skews the distribution even more. For that reason, it was not transformed, which is why the log-transformed score is not included in Table 17 and the standardisation on a 0-to-1 scale was calculated directly on the score.

Figure 6: Distribution of countries according to the mean and standard deviation of the share of public/private co-publications normalised by the world weighted average and the log-transformed equivalent (2019)



Source: These statistics were calculated using data from Scopus (Elsevier)

Table 18 shows the SPP relative to the world weighted average for each KA. The EU27's SPP was well above the world weighted average across all KAs. The MIMs' SPP was also above the world weighted average in every KA but much less so. As was mentioned previously, some countries failed to meet the minimum threshold of 30 publications in a few KAs, for which no score was calculated. As was the case for all other indicators, the same countries tended to rank highly across all KAs, with few exceptions. The EU27's SPP increased by 3 % or more annually in KA2, KA4, KA5 and KA7, but it decreased by 3.5 % in KA3. The MIMs' performance was more stable, fluctuating by less than 1 % annually in every KA, except KA1 (1.2 % annual increase) and KA3 (1.1 % annual increase).

Table 18: Share of public/private co-publications normalised by the world weighted average for each KA (2019) with corresponding trendline (2015-2019)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EU27	**EU27**	**1.41**	**1.42**	**1.58**	**1.19**	**1.29**	**1.50**	**1.55**	**1.72**	**1.25**	**1.23**
MIM	**MIM**	**1.09**	**1.12**	**1.03**	**1.12**	**1.13**	**1.14**	**1.03**	**1.09**	**1.05**	**1.05**
Austria	AT	2.55	2.93	2.58	1.89	2.03	2.21	2.97	5.12	1.77	1.34
Netherlands	NL	1.95	2.17	2.41	1.57	1.51	1.27	2.43	2.78	1.65	1.42
Croatia	HR	1.92	1.99	2.35	1.84	N/C	N/C	1.15	2.61	N/C	N/C
Slovenia	SI	1.84	1.60	1.88	1.34	2.00	N/C	2.51	3.30	N/C	1.20
Germany	DE	1.74	1.79	1.88	1.46	2.06	2.40	1.92	2.13	1.48	1.13
Finland	FI	1.72	1.33	1.86	1.18	1.81	1.72	0.78	2.39	0.98	2.32
Sweden	SE	1.65	1.34	1.39	1.43	1.76	2.22	2.04	2.46	1.68	1.49
Luxembourg	LU	1.59	N/C	2.23	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Ireland	IE	1.56	1.58	1.81	1.38	1.42	1.46	1.81	1.87	1.06	N/C
Japan	JP	1.56	1.59	1.68	1.42	2.22	2.14	1.56	1.79	1.25	1.20
Denmark	DK	1.52	1.70	1.22	1.08	1.97	2.00	1.48	2.43	1.85	0.82
Norway	NO	1.49	1.76	1.72	0.96	1.61	1.87	1.10	1.36	1.84	1.96
Belgium	BE	1.49	1.13	1.43	1.57	2.13	1.90	1.69	1.79	1.55	1.06
France	FR	1.46	1.41	1.62	1.23	1.51	1.03	1.51	1.44	1.57	1.32
Greece	EL	1.43	1.46	1.98	1.13	1.73	1.65	1.10	1.64	0.54	N/C
Cyprus	CY	1.42	1.88	1.95	1.02	0.97	N/C	N/C	N/C	N/C	N/C
Italy	IT	1.34	1.22	1.53	1.38	1.10	1.37	1.61	1.49	1.24	0.99
Estonia	EE	1.30	0.84	1.71	0.96	1.34	N/C	N/C	0.79	N/C	N/C
Hungary	HU	1.28	1.60	0.84	1.12	0.71	N/C	0.00	2.12	N/C	1.39
United Kingdom	UK	1.27	1.24	1.19	0.97	1.09	1.05	1.47	1.73	1.23	1.60
United States	US	1.17	1.12	1.24	0.98	1.17	1.14	1.03	1.37	1.30	1.26
China	CN	1.15	1.32	1.02	1.43	1.05	0.95	0.90	0.79	0.86	0.72
Canada	CA	1.13	1.24	0.96	1.04	0.91	1.17	1.29	1.48	0.98	1.20
Spain	ES	1.04	1.15	1.40	0.78	1.03	0.70	1.25	0.86	0.58	1.47
Rep. of Korea	KR	0.95	0.88	1.10	0.82	1.20	0.86	0.83	1.22	0.92	0.89
Czech Republic	CZ	0.84	1.04	0.66	0.34	0.84	1.15	0.82	0.93	1.23	1.03
Lithuania	LT	0.80	1.33	0.90	N/C	N/C	N/C	N/C	0.99	N/C	N/C
Slovakia	SK	0.80	0.40	1.21	0.65	0.26	1.04	0.12	1.01	N/C	1.67
Portugal	PT	0.75	0.77	0.93	0.75	0.52	0.90	0.68	0.78	0.69	0.64
Romania	RO	0.74	0.72	0.95	0.90	0.30	0.23	1.21	0.61	0.35	0.28
Poland	PL	0.70	0.66	1.21	0.57	0.79	0.53	0.44	0.76	0.54	0.86
Bulgaria	BG	0.62	0.25	0.90	0.74	N/C	N/C	0.00	1.29	N/C	N/C
Australia	AU	0.61	0.65	0.58	0.52	0.54	0.73	0.44	0.87	0.94	0.64
Brazil	BR	0.57	0.54	0.50	0.66	0.19	0.55	0.68	0.59	1.16	0.89
United Arab Emirates	AE	0.49	0.66	0.78	0.38	0.44	0.33	0.00	0.28	0.63	N/C
Chile	CL	0.49	0.38	0.44	0.49	0.24	N/C	0.44	1.06	N/C	N/C
Latvia	LV	0.43	N/C	0.36	0.32	0.58	N/C	N/C	0.33	N/C	N/C
Saudi Arabia	SA	0.43	0.33	0.42	0.44	0.29	0.72	0.64	0.57	0.93	0.18
Mexico	MX	0.31	0.20	0.66	0.21	0.14	0.49	0.52	0.47	0.18	0.20
India	IN	0.26	0.19	0.31	0.21	0.46	0.44	0.32	0.33	0.32	0.26
Indonesia	ID	0.24	0.22	0.19	0.48	0.20	0.09	0.36	0.29	0.42	0.00
Malta	MT	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier)

3.9 Share of highly cited publications among the 10 % most cited

Box 3-9: Conclusion regarding the share of highly cited publications among the 10 % most cited for inclusion in the CEII

> Following these analyses, even though some countries did not meet the 30 publications threshold in some KAs, we conclude that the share of highly cited publications among the 10 % most cited is a pertinent indicator to measure the level of research impact in clean energy research, one of the scientific aspects of interest to the Commission, and to be included in the CEII. The share of highly cited publications is a normalised indicator that enables comparison between countries. The outlier and break in series tests identified a few potentially problematic data points, which were investigated but were ultimately not a cause for concern. Finally, this indicator can easily be transformed to be included in the CEII.

The share of highly cited publications among the 10 % most cited ($HCP_{10\%}$) is calculated based on the number of publications among the 10 % most cited and the total number of publications in clean energy research. The share of $HCP_{10\%}$ is shown in Table 19 for 2017 and all KAs combined. The "score" column is the $HCP_{10\%}$ (i.e. share of papers in the 10 % most cited papers in Scopus) multiplied by 10 to reflect an expected score centred at 1; however, the RC scores of the publications in the KA data sets are based on the whole of Scopus and were not renormalised by KA, instead relying on the normalisation according to the 174 subfields of Science-Metrix's classification of science. Therefore, an $HCP_{10\%}$ of 1.52 at the world level indicates that clean energy publications include 52 % (5.2 p.p.) more highly cited publications than expected. Since the world level varies from 0.44 to 1.95 across KAs (only one KA is markedly below expectations, see Table 20), it is best to compare a country's performance across KAs using the re-scaled scores (between 0 and 1); this provides an assessment of performance, relative to other countries, across KAs rather than on the absolute score for $HCP_{10\%}$. A minimum of 30 publications was also applied to the $HCP_{10\%}$ given that a small share of publications are among the 10 % most cited, which again only affected Malta for all KAs combined.

Looking more closely at the results in Table 19, the EU27's collective impact was slightly below the world level, at 14.6 %, while the MIMs' impact was slightly above, at 15.8 %. Two of the EU27 members (Denmark and Greece) figure in Cluster 1, and all countries but one in Cluster 4 are EU27 members. In fact, 16 members of the EU27 obtained a score below the world level. Many MIMs obtained high scores, such as Australia at the top, as well as Saudi Arabia and the United Kingdom, with more than 20 % of their clean energy publications being highly cited.

The number of publications among the 10 % most cited and the total number of publications were both computed using a fractional count. A reminder that only publications with an RC score are included, which excludes conference papers. The publications that figure among the 10 % most cited are part of an elite group. At the country and KA levels, this results in a small number of publications, even for countries with a large output, which is compounded by the fact that a fractional count is used. Nevertheless, while disregarding countries that did not meet the 30-publication threshold, few flags were raised by the outlier and break in time series tests run on the number of publications among the 10 % most cited, and they were investigated and deemed appropriate. The tests were also run on the total number of publications without conference papers, in fractional counting, which also often resulted in small numbers at the country and KA levels, but no potential outliers were flagged. The test run on the share of $HCP_{10\%}$ flagged a few potential outliers: the whole of MIMs in KA2 in 2017 and the whole of the EU27 members in KA3 in 2013, which did not excessively diverge away from the overall trend.

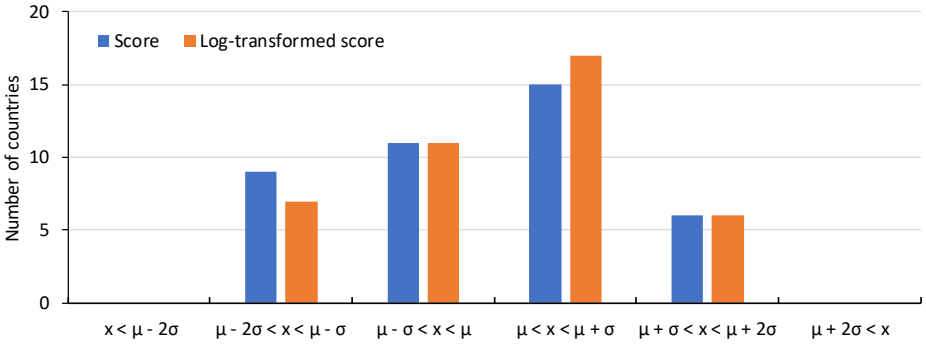
Table 19: Share of highly cited publications among the 10 % most cited for all KAs combined

Country	Code	Share in GDP (2019)	$HCP_{10\%}$ (2017)	Score (2017)	Score 0-1 (2017)	CAGR		
						2013 to 2017	Gap to EU27	Gap to MIM
World	**World**	**100%**	**15.2%**	**1.52**	**0.52**	**-0.4%**	**2.4%**	**-0.4%**
EU27	**EU27**	**15.3%**	**14.6%**	**1.46**	**0.49**	**-2.8%**	**0.0%**	**-2.8%**
MIM	**MIM**	**69.3%**	**15.8%**	**1.58**	**0.54**	**0.0%**	**2.8%**	**0.0%**
Cluster 1		**20.9%**						
Australia	AU	1.0%	24.2%	2.42	0.84	3.1%	5.9%	3.1%
Denmark	DK	0.3%	22.1%	2.21	0.77	-5.6%	-2.8%	-5.6%
Saudi Arabia	SA	1.2%	21.5%	2.15	0.75	-0.9%	1.9%	-0.9%
Greece	EL	0.2%	21.2%	2.12	0.73	4.5%	7.3%	4.5%
United Kingdom	UK	2.4%	20.7%	2.07	0.72	-1.0%	1.8%	-1.0%
United States	US	15.8%	18.9%	1.89	0.65	-6.0%	-3.2%	-6.0%
Cluster 2		**29.1%**						
Italy	IT	2.0%	18.5%	1.85	0.64	2.0%	4.8%	2.0%
Netherlands	NL	0.8%	18.4%	1.84	0.64	-6.8%	-4.0%	-6.8%
United Arab Emirates	AE	0.5%	18.4%	1.84	0.63	3.9%	6.7%	3.9%
Ireland	IE	0.3%	16.8%	1.68	0.58	-5.6%	-2.8%	-5.6%
China	CN	17.7%	16.8%	1.68	0.57	7.0%	9.8%	7.0%
Portugal	PT	0.3%	16.7%	1.67	0.57	1.9%	4.7%	1.9%
Belgium	BE	0.5%	16.7%	1.67	0.57	-8.6%	-5.8%	-8.6%
Canada	CA	1.4%	16.6%	1.66	0.57	-3.5%	-0.7%	-3.5%
Estonia	EE	0.0%	16.4%	1.64	0.56	19.6%	22.4%	19.6%
Sweden	SE	0.4%	16.2%	1.62	0.55	-5.1%	-2.3%	-5.1%
Luxembourg	LU	0.1%	15.7%	1.57	0.54	-8.4%	-5.6%	-8.4%
Finland	FI	0.2%	14.9%	1.49	0.51	1.0%	3.8%	1.0%
Germany	DE	3.4%	14.7%	1.47	0.50	-4.2%	-1.4%	-4.2%
Cyprus	CY	0.0%	14.4%	1.44	0.49	-15.7%	-12.9%	-15.7%
Spain	ES	1.5%	14.2%	1.42	0.48	-5.2%	-2.4%	-5.2%
Cluster 3		**21.8%**						
Chile	CL	0.4%	12.7%	1.27	0.43	-10.8%	-8.0%	-10.8%
Rep. of Korea	KR	1.7%	12.3%	1.23	0.41	1.1%	3.9%	1.1%
France	FR	2.4%	12.1%	1.21	0.40	-5.1%	-2.3%	-5.1%
Norway	NO	0.3%	12.0%	1.20	0.40	-8.3%	-5.5%	-8.3%
India	IN	7.1%	11.2%	1.12	0.37	-0.8%	2.0%	-0.8%
Austria	AT	0.4%	10.5%	1.05	0.35	-6.5%	-3.7%	-6.5%
Japan	JP	4.0%	8.6%	0.86	0.28	-5.0%	-2.2%	-5.0%
Croatia	HR	0.1%	8.2%	0.82	0.27	-4.9%	-2.1%	-4.9%
Poland	PL	1.0%	7.9%	0.79	0.25	12.1%	14.9%	12.1%
Brazil	BR	2.4%	7.7%	0.77	0.25	-0.9%	1.9%	-0.9%
Hungary	HU	0.2%	7.5%	0.75	0.24	6.2%	9.0%	6.2%
Mexico	MX	1.9%	7.2%	0.72	0.23	5.8%	8.6%	5.8%
Cluster 4		**3.7%**						
Romania	RO	0.4%	7.2%	0.72	0.23	12.7%	15.5%	12.7%
Slovenia	SI	0.1%	6.4%	0.64	0.20	-4.8%	-2.0%	-4.8%
Czech Republic	CZ	0.3%	5.2%	0.52	0.16	5.9%	8.7%	5.9%
Indonesia	ID	2.5%	4.6%	0.46	0.14	-22.4%	-19.6%	-22.4%
Lithuania	LT	0.1%	4.5%	0.45	0.13	8.5%	11.3%	8.5%
Bulgaria	BG	0.1%	4.1%	0.41	0.12	0.7%	3.5%	0.7%
Slovakia	SK	0.1%	4.1%	0.41	0.12	-4.4%	-1.6%	-4.4%
Latvia	LV	0.0%	4.0%	0.40	0.12	6.6%	9.4%	6.6%
Not calculated		**0.0%**						
Malta	MT	0.0%	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

The EU27's and MIMs' shares of $HCP_{10\%}$ were near normally distributed, as can be seen in Figure 7. The logarithm skews the distribution toward above-average countries; transforming the $HCP_{10\%}$ was therefore unnecessary, which is why it is not included in Table 19. The score was used to standardise this indicator on a 0-to-1 scale.

Figure 7: Distribution of countries according to the mean and standard deviation of the share of highly cited publications among the 10 % most cited and the log-transformed equivalent (2017)



Source: These statistics were calculated using data from Scopus (Elsevier)

Table 20 shows the $HCP_{10\%}$ for each KA. All KAs were hot topics that included a substantial share of their publications among the most cited, as can be seen in the world scores well above 1 in every KA, except KA9. The EU27 performed substantially better than the world in KA2, KA4 and KA9. The MIMs performed better than the world in all KAs but only stood out in KA4 and KA5. The countries that performed well tended to do so across all KAs. Australia and Denmark performed quite well in almost all KAs. Many countries in some KAs did not publish enough articles for the score to be calculated, which tended to be lower ranking countries. The EU27's shares of $HCP_{10\%}$ decreased substantially in KA1 (by 5.1 % annually), KA6 (by 9.0 % annually) and KA8 (by 7.0 % annually), while it increased only in KA2 (by 2.5 % annually) and KA5 (by 3.9 % annually). The MIMs' performance was more mixed, increasing by about 4 % or more in KA2 and KA5, and decreasing by as much in KA6 and KA8.

Table 20: Share of highly cited publications among the 10 % most cited for each KA (2017) with corresponding trendline (2013-2017)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.52**	**1.42**	**1.66**	**1.72**	**1.46**	**1.53**	**1.95**	**1.55**	**1.71**	**0.44**
EU27	**EU27**	**1.46**	**1.41**	**1.90**	**1.70**	**1.59**	**1.55**	**1.56**	**1.37**	**1.65**	**0.57**
MIM	**MIM**	**1.58**	**1.46**	**1.71**	**1.78**	**1.57**	**1.68**	**2.03**	**1.62**	**1.74**	**0.49**
Australia	AU	2.42	2.37	2.75	2.88	2.57	1.78	3.68	2.02	1.81	0.80
Denmark	DK	2.21	1.74	3.17	3.65	3.02	2.42	2.83	1.37	1.56	N/C
Saudi Arabia	SA	2.15	1.93	2.19	3.63	1.95	2.02	2.38	1.94	1.48	N/C
Greece	EL	2.12	2.48	2.59	1.92	1.78	N/C	3.45	1.64	3.03	N/C
United Kingdom	UK	2.07	2.01	2.71	2.42	2.12	2.24	2.69	1.94	1.86	0.86
United States	US	1.89	2.05	2.14	2.33	1.63	1.80	2.82	1.55	2.02	0.76
Italy	IT	1.85	1.79	2.41	2.08	2.38	2.19	2.00	1.63	1.56	0.53
Netherlands	NL	1.84	1.88	1.65	2.26	2.00	1.67	2.35	1.76	1.63	0.69
United Arab Emirates	AE	1.84	1.63	1.40	2.28	2.46	N/C	N/C	1.74	1.16	N/C
Ireland	IE	1.68	1.24	2.30	2.13	2.63	N/C	N/C	0.98	N/C	N/C
China	CN	1.68	1.39	1.61	1.71	1.53	2.04	2.01	2.12	1.88	0.33
Portugal	PT	1.67	1.82	1.93	1.78	1.57	0.88	1.75	1.99	1.34	0.36
Belgium	BE	1.67	1.45	2.66	1.70	1.86	1.80	1.27	2.08	2.33	0.30
Canada	CA	1.66	1.40	2.26	2.25	1.32	1.02	2.88	1.18	1.58	0.40
Estonia	EE	1.64	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Sweden	SE	1.62	1.32	3.36	2.20	1.56	1.47	1.43	1.43	1.13	0.65
Luxembourg	LU	1.57	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Finland	FI	1.49	2.13	2.58	1.63	1.19	1.41	N/C	0.98	0.80	0.45
Germany	DE	1.47	1.46	1.32	1.62	1.39	1.39	1.88	1.48	1.74	1.03
Cyprus	CY	1.44	1.15	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Spain	ES	1.42	1.31	1.83	1.67	1.38	1.26	1.60	1.28	1.92	0.36
Chile	CL	1.27	1.58	N/C	1.57	1.16	N/C	N/C	0.90	N/C	N/C
Rep. of Korea	KR	1.23	1.16	1.27	1.67	0.50	0.90	1.54	1.30	1.34	0.11
France	FR	1.21	1.07	1.92	1.51	0.57	1.37	1.16	1.50	1.95	0.45
Norway	NO	1.20	1.08	2.06	1.75	1.41	N/C	2.66	0.71	0.54	N/C
India	IN	1.12	1.05	1.22	1.25	0.91	1.08	0.92	1.21	1.42	0.30
Austria	AT	1.05	1.04	1.52	0.75	0.41	1.59	0.41	1.25	1.08	0.67
Japan	JP	0.86	0.92	1.28	0.73	0.95	0.71	0.80	0.97	1.33	0.35
Croatia	HR	0.82	0.24	N/C	0.56	N/C	N/C	N/C	N/C	N/C	N/C
Poland	PL	0.79	0.68	1.02	0.79	0.93	1.38	0.85	0.74	0.91	0.07
Brazil	BR	0.77	0.54	1.43	0.87	0.87	1.02	2.83	0.59	1.19	0.01
Hungary	HU	0.75	1.26	0.68	0.80	N/C	N/C	N/C	0.73	N/C	0.25
Mexico	MX	0.72	0.73	0.72	0.69	N/C	0.25	N/C	0.70	1.21	N/C
Romania	RO	0.72	0.35	1.36	1.13	0.25	N/C	N/C	1.12	N/C	0.12
Slovenia	SI	0.64	0.28	N/C	N/C	N/C	N/C	N/C	0.54	N/C	0.11
Czech Republic	CZ	0.52	0.52	0.57	0.30	0.30	N/C	0.01	0.69	1.56	0.23
Indonesia	ID	0.46	0.50	0.17	0.67	N/C	N/C	N/C	0.58	N/C	N/C
Lithuania	LT	0.45	0.90	0.59	N/C	N/C	N/C	N/C	0.00	N/C	N/C
Bulgaria	BG	0.41	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Slovakia	SK	0.41	N/C	N/C	0.00	N/C	N/C	N/C	0.70	N/C	0.01
Latvia	LV	0.40	N/C	N/C	N/C	N/C	N/C	N/C	0.73	N/C	N/C
Malta	MT	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier)

3.10 Share of publications cited by patents

Box 3-10: Conclusion regarding the share of publications cited by patents for inclusion in the CEII

> The SCP is a pertinent indicator to measure the level of knowledge transfer to innovation in clean energy research, one of the scientific aspects of interest to the Commission, especially given the fact that it measures the translation of scientific research into innovation. However, citations of publications by patents are quite rare, resulting in many EU27 members and MIMs in many KAs with none of their publications cited by a single patent. This makes the comparison of the members' performance impractical. For this reason, we do not recommend including it in the CEII. Additionally, the dimension tracked by this indicator is already well captured by the SPP.

The share of publication cited by patents (SCP) is calculated based on the number of publications cited by a patent and the total number of publications in clean energy research. The SCP is shown in Table 21 for 2015 and all KAs combined and is normalised according to the world weighted average (the "score" column). A fairly small share of 3.2 % of clean energy publications were cited by a patent in 2015 globally, which is why a minimum of 30 publications was imposed to calculate this indicator for a specific country or aggregation of countries. Once again, Malta was the only country that did not meet this threshold. The SCP was slightly lower for EU27 members, with 2.7 %, and slightly higher for MIMs, with 3.5 %. One EU27 member, Luxembourg, placed in Cluster 1, whereas all countries in Cluster 4, except for Indonesia, were EU27 members. The EU27's SCP normalised by the world weighted average decreased by 1.3 % annually between 2011 and 2015, but the trends at the national level were very volatile given the nature of this indicator. None of Bulgaria's publications in 2015, in any KA, were cited by a single patent.

The number of publications cited by a patent and the total number of publications are both computed using a full count at country level but fractioned between KAs. The trend of the total number of publications was already tested for other indicators, and the tests are therefore not repeated here. The outlier and break in time series tests were applied on the number of publications cited by a patent, and the SCP normalised by the world weighted average. As can be seen in Table 21, only very few clean energy publications were cited by a patent. At the country and KA levels, the numbers are quite small and tended to fluctuate, even for countries with a substantial output of publications. The tests identified some potential outliers and breaks in time series, all of which were investigated and deemed accurate given the small number of cited publications. The outlier test run on the SCP itself flagged quite a few potential outliers in 2010, followed by a sharp decline during the following years. The year 2010 is outside the period of interest but this is still worth mentioning because such a behaviour is actually expected for this indicator. The number of publications tends to increase over time, but publications at the beginning of the time series have had the chance to be cited for 10 years, and naturally more of them had been cited than articles published recently. As a result, the SCP naturally decreases over time, which is why it is normalised by the world weighted average in each corresponding year. The tests run on the SCP normalised by the world weighted average identified very few potential outliers and breaks in time series between 2011 and 2015, which were ultimately rationalised by the small number of publications cited by patents.

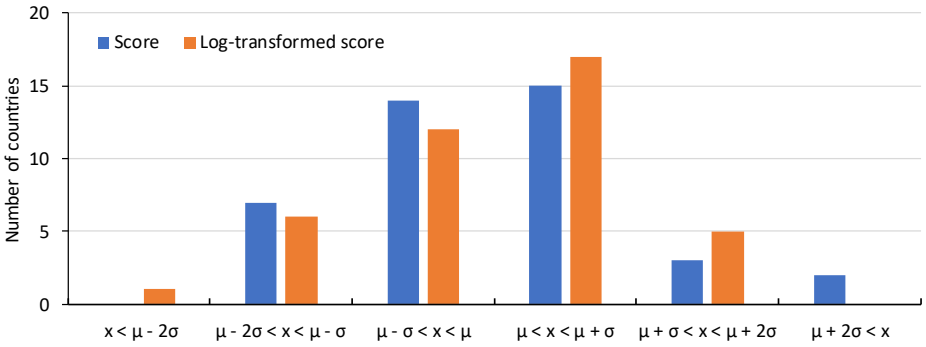
The SCP exhibits a fairly "normal-looking" distribution (**Error! Reference source not found.**). Applying the logarithm to the SCP heavily skews the distribution in favour of above-average countries; it was therefore decided not to transform the SCP and to standardise the score directly on a 0-to-1 scale.

Table 21: Share of publications cited by patents for all KAs combined

Country	Code	Share in GDP (2019)	SCP (2015)	Score (2015)	Score 0-1 (2015)	CAGR		
						2011 to 2015	Gap to EU27	Gap to MIM
World	**World**	**100%**	**3.2%**	**1.00**	**0.55**	**0.0%**	**1.3%**	**-0.5%**
EU27	**EU27**	**15.3%**	**2.7%**	**0.85**	**0.47**	**-1.3%**	**0.0%**	**-1.8%**
MIM	**MIM**	**69.3%**	**3.5%**	**1.08**	**0.60**	**0.5%**	**1.8%**	**0.0%**
Cluster 1		**20.2%**						
Saudi Arabia	SA	1.2%	5.7%	1.78	0.98	13.0%	14.3%	12.5%
United States	US	15.8%	5.4%	1.70	0.93	4.0%	5.3%	3.5%
Luxembourg	LU	0.1%	5.1%	1.58	0.87	4.0%	5.3%	3.5%
Rep. of Korea	KR	1.7%	4.9%	1.52	0.84	2.7%	4.0%	2.2%
Canada	CA	1.4%	4.3%	1.34	0.74	2.8%	4.1%	2.3%
Cluster 2		**35.4%**						
United Kingdom	UK	2.4%	3.8%	1.19	0.66	5.4%	6.7%	4.9%
Australia	AU	1.0%	3.7%	1.16	0.64	3.6%	4.9%	3.1%
Germany	DE	3.4%	3.7%	1.14	0.63	1.7%	3.0%	1.2%
Greece	EL	0.2%	3.4%	1.07	0.59	3.3%	4.6%	2.8%
Netherlands	NL	0.8%	3.3%	1.04	0.57	-5.3%	-4.0%	-5.8%
Spain	ES	1.5%	3.2%	1.01	0.56	1.7%	3.0%	1.2%
Finland	FI	0.2%	3.2%	1.01	0.56	4.9%	6.2%	4.4%
Denmark	DK	0.3%	3.2%	1.00	0.55	-6.4%	-5.1%	-6.9%
Belgium	BE	0.5%	3.2%	0.99	0.54	1.4%	2.7%	0.9%
Japan	JP	4.0%	3.1%	0.95	0.53	-1.6%	-0.3%	-2.1%
Sweden	SE	0.4%	3.1%	0.95	0.52	-3.0%	-1.7%	-3.5%
China	CN	17.7%	2.8%	0.88	0.49	-0.5%	0.8%	-1.0%
Austria	AT	0.4%	2.8%	0.86	0.48	-4.1%	-2.8%	-4.6%
France	FR	2.4%	2.7%	0.84	0.46	-0.2%	1.1%	-0.7%
Norway	NO	0.3%	2.5%	0.79	0.43	3.2%	4.5%	2.7%
Cluster 3		**16.3%**						
Slovenia	SI	0.1%	2.5%	0.77	0.42	11.5%	12.8%	11.0%
Mexico	MX	1.9%	2.4%	0.74	0.41	3.4%	4.7%	2.9%
Italy	IT	2.0%	2.2%	0.70	0.38	-2.0%	-0.7%	-2.5%
Chile	CL	0.4%	2.2%	0.68	0.38	-17.2%	-15.9%	-17.7%
United Arab Emirates	AE	0.5%	2.1%	0.66	0.37	-1.3%	0.0%	-1.8%
Croatia	HR	0.1%	1.9%	0.58	0.32	-3.1%	-1.8%	-3.6%
Ireland	IE	0.3%	1.8%	0.57	0.31	-18.3%	-17.0%	-18.8%
Estonia	EE	0.0%	1.7%	0.54	0.30	-21.5%	-20.2%	-22.0%
Lithuania	LT	0.1%	1.6%	0.49	0.27	4.9%	6.2%	4.4%
India	IN	7.1%	1.5%	0.47	0.26	-3.6%	-2.3%	-4.1%
Portugal	PT	0.3%	1.4%	0.45	0.25	-8.5%	-7.2%	-9.0%
Hungary	HU	0.2%	1.4%	0.42	0.23	-3.7%	-2.4%	-4.2%
Brazil	BR	2.4%	1.2%	0.38	0.21	-6.7%	-5.4%	-7.2%
Poland	PL	1.0%	1.2%	0.36	0.20	0.3%	1.6%	-0.2%
Cluster 4		**3.6%**						
Slovakia	SK	0.1%	1.1%	0.33	0.18	11.7%	13.0%	11.2%
Cyprus	CY	0.0%	1.0%	0.31	0.17	-32.0%	-30.7%	-32.5%
Romania	RO	0.4%	0.8%	0.24	0.13	-10.5%	-9.2%	-11.0%
Czech Republic	CZ	0.3%	0.7%	0.21	0.12	-15.6%	-14.3%	-16.1%
Indonesia	ID	2.5%	0.6%	0.17	0.10	-5.6%	-4.3%	-6.1%
Latvia	LV	0.0%	0.4%	0.13	0.07	-30.0%	-28.7%	-30.5%
Bulgaria	BG	0.1%	0.0%	0.00	0.00	-100.0%	-98.7%	-100.5%
Not calculated		**0.0%**						
Malta	MT	0.0%	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Figure 8: Distribution of countries according to the mean and standard deviation of the share of publications cited by patents normalised by the world weighted average and the log-transformed equivalent (2015)



Source: These statistics were calculated using data from Scopus (Elsevier)

Table 22 shows the SCP normalised by the world weighted average for each KA. The EU27's SCP was above the world weighted average in a single KA (KA5), but marginally more so. The MIMs performed slightly better than the world weighted average in every KA. In some KAs, many countries did not publish enough publications to register a score, which is compounded by the many countries in some KAs with a score of 0, meaning that none of their publications in these KAs in 2015 were cited by a single patent. These countries tended to be lower ranking, meaning once again that the countries that performed well tended to do so across all KAs. KA9 appears to be somewhat of an oddity, with only 15 countries having the sufficient number of publications and at least one of them cited by a patent. Some scientific subfields align very well with patenting activities; nuclear safety (KA9) does not appear to be one of them. The EU27's SCP decreased in every KA, particularly in KA3 (by 4.4 % annually) and KA4 (by 7.3 % annually). At the opposite, the MIMs' performance increased in every KA (except KA4), but by less than 3 % annually.

Table 22: Share of publications cited by patents normalised by the world weighted average for each KA (2015) with corresponding trendline (2011-2015)

Country	Code	All KAs	KA1	KA2	KA3	KA4	KA5	KA6	KA7	KA8	KA9
World	**World**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
EU27	**EU27**	**0.85**	**0.94**	**0.95**	**0.87**	**0.73**	**1.10**	**0.67**	**0.97**	**0.91**	**0.69**
MIM	**MIM**	**1.08**	**1.09**	**1.09**	**1.08**	**1.09**	**1.15**	**1.05**	**1.05**	**1.04**	**1.07**
Saudi Arabia	SA	1.78	1.91	1.26	1.33	1.64	0.85	0.00	1.60	2.72	N/C
United States	US	1.70	1.79	1.72	1.93	1.77	1.38	1.69	1.54	1.46	2.44
Luxembourg	LU	1.58	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Rep. of Korea	KR	1.52	1.43	1.40	1.37	1.93	1.53	1.67	1.17	1.82	1.21
Canada	CA	1.34	1.23	1.18	1.48	1.05	1.70	1.53	1.36	1.03	0.00
United Kingdom	UK	1.19	1.14	1.66	1.10	0.61	2.06	0.92	1.46	0.82	1.32
Australia	AU	1.16	1.29	0.92	1.33	1.50	1.40	0.84	1.03	0.80	1.37
Germany	DE	1.14	1.16	1.36	1.04	0.85	1.41	0.96	1.25	1.45	1.14
Greece	EL	1.07	1.19	0.78	0.82	1.27	N/C	0.00	1.61	0.79	0.00
Netherlands	NL	1.04	1.33	0.47	0.60	1.66	0.00	0.40	1.92	0.66	1.43
Spain	ES	1.01	1.07	0.97	1.35	0.49	1.43	1.05	1.04	0.72	0.00
Finland	FI	1.01	1.34	1.79	0.47	0.00	1.41	0.00	0.76	0.00	2.78
Denmark	DK	1.00	0.71	0.44	1.27	0.00	1.43	0.58	1.52	1.49	0.00
Belgium	BE	0.99	0.93	1.45	1.09	0.79	0.75	0.65	1.38	0.74	2.10
Japan	JP	0.95	1.04	1.20	0.73	1.82	0.57	1.08	1.01	1.07	0.51
Sweden	SE	0.95	0.61	1.29	1.49	0.39	1.68	0.83	1.05	0.80	0.96
China	CN	0.88	0.88	0.80	0.88	0.88	0.77	0.84	0.85	0.67	0.50
Austria	AT	0.86	1.43	0.70	1.35	0.00	1.63	0.24	0.64	0.78	2.66
France	FR	0.84	1.02	1.05	1.00	1.57	1.33	0.61	1.03	1.03	0.21
Norway	NO	0.79	0.34	1.02	0.97	1.46	2.55	0.25	1.10	0.95	0.00
Slovenia	SI	0.77	0.73	1.11	0.00	N/C	N/C	N/C	0.00	N/C	0.00
Mexico	MX	0.74	0.56	0.55	1.15	N/C	1.40	N/C	0.75	0.66	0.00
Italy	IT	0.70	0.98	0.84	0.51	0.88	0.93	0.54	0.69	0.59	0.00
Chile	CL	0.68	0.36	N/C	2.14	0.00	N/C	N/C	0.38	N/C	N/C
United Arab Emirates	AE	0.66	0.41	1.20	0.44	0.00	N/C	N/C	0.49	0.76	N/C
Croatia	HR	0.58	0.54	1.13	0.79	N/C	N/C	N/C	N/C	N/C	N/C
Ireland	IE	0.57	0.79	0.62	0.00	0.97	N/C	N/C	0.00	N/C	N/C
Estonia	EE	0.54	0.00	N/C	0.00	N/C	N/C	N/C	0.73	N/C	N/C
Lithuania	LT	0.49	1.89	0.00	N/C	0.00	N/C	N/C	0.00	N/C	0.00
India	IN	0.47	0.33	0.39	0.48	0.98	0.30	0.47	0.57	0.83	1.16
Portugal	PT	0.45	0.39	0.40	0.97	0.41	0.00	0.15	0.45	0.76	0.00
Hungary	HU	0.42	0.29	0.61	1.19	N/C	N/C	N/C	0.69	N/C	0.00
Brazil	BR	0.38	0.00	0.43	0.39	0.00	0.00	0.78	0.47	0.67	0.00
Poland	PL	0.36	0.25	0.34	0.23	1.79	0.00	0.73	0.46	0.29	0.00
Slovakia	SK	0.33	0.00	0.45	0.37	0.00	N/C	N/C	0.55	N/C	0.00
Cyprus	CY	0.31	0.90	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C
Romania	RO	0.24	0.16	0.21	0.22	0.00	0.00	0.32	0.29	0.00	3.69
Czech Republic	CZ	0.21	0.22	0.23	0.50	0.00	0.00	0.00	0.18	N/C	0.00
Indonesia	ID	0.17	0.22	0.16	0.15	0.00	N/C	0.26	0.16	N/C	N/C
Latvia	LV	0.13	0.00	0.00	0.00	0.00	N/C	N/C	0.52	N/C	N/C
Bulgaria	BG	0.00	N/C	N/C	N/C	N/C	N/C	N/C	0.00	N/C	N/C
Malta	MT	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C	N/C

Source: These statistics were calculated using data from Scopus (Elsevier)

4 Conclusion

Science-Metrix proposed 10 bibliometric indicators as candidates for inclusion in the CEII. The indicators are designed to measure the following scientific aspects, which are of high relevance to the policy context surrounding the development of the CEII (i.e. the SET Plan, the Energy Union, and the Mission Innovation initiative): output, collaboration, open access and impact. All the indicators were calculated for each EU27 member and MIM, except for two indicators that were designed specifically for EU27 members. This report presented the results for each of these indicators and conducted a number of tests to assess the quality of the data that feed into them and their soundness for inclusion in the CEII.

It is important to mention that no indicator is flawless. Section 2.1 highlighted many of the shortcomings associated with publication-based indicators. A composite indicator, such as the CEII, can nevertheless be reliable even if it includes many flawed indicators, as long as it adheres to the method known as "converging partial indicators"[6,7]. This method argues that the flaws in the indicators meant to evaluate scientific performance can be overlooked if the following conditions are met: (1) the indicators are applied to research groups rather than individual scientists; (2) the indicators based on citations are seen as reflecting the impact, rather than the quality or importance, of the research work; (3) a range of indicators are employed, each of which focuses on different aspects of a group's performance; (4) the indicators are applied to matched groups; and (5) they yield convergent results. In this case, all conditions are met.

The indicators considered that measure scientific output are the following:

- Number of publications
- Specialisation index

The number of publications was deemed appropriate for potential inclusion in the CEII; however, it needs to be normalised by a metric that considers the size of country since larger countries naturally tend to publish more articles. We proposed normalising the number of publications by the population data. The SI is by definition a normalised index and was deemed appropriate for consideration in the CEII.

The indicators that measure scientific collaboration are the following:

- Share of international co-publications
- Share of transnational co-publications
- Weighted eigenvector centrality in the world's co-publication network
- Weighted eigenvector centrality in the Member States' co-publication network
- Share of public/private co-publications

The STP and the WEC in the Member States' co-publication network were calculated only for EU27 members. The SIP and the SPP were deemed appropriate for inclusion in the CEII, contrary to the STP and the WEC calculated for both networks. The STP was found

[6] Martin, B.R. and Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. Research Policy, 12, 61–90.

[7] Martin, B.R. (1996). The use of multiple indicators in the assessment of basic research. Scientometrics, 36, 343–362.

to be strongly correlated with the SIP; therefore, it adds little value beyond the SIP, and it is not calculated for MIMs. The WEC is also strongly correlated with the number of publications, regardless of the network, and the countries that published the most articles therefore naturally had the highest WEC.

The share of open access publications is the only indicator that measures scientific open access. We concluded that it is an appropriate indicator worth considering for inclusion in the CEII.

Finally, the indicators that measure scientific impact and knowledge transfer in innovation are the following:

- Share of publications among the 10 % most cited (i.e. highly cited publications or $HCP_{10\%}$)
- Share of publications cited in patents

The share of $HCP_{10\%}$ was deemed appropriate for consideration in the CEII, but not the SCP. The SCP is a publication-to-patent citation indicator. In theory, measuring article citations in patents is quite valuable since it informs us on the translation of scientific research into innovation. In practice, however, so few articles are ever cited by patents that many EU27 members and MIMs, in many KAs, have none of their publications cited by a single patent, which makes a comparison between them impractical.

As a last analysis, Table 23 shows the correlation coefficient between the indicators common to both EU27 members and MIMs. As was mentioned previously, the correlation between the number of publications and the WEC in the world's co-publication network was high, suggesting that the WEC is not an appropriate indicator to compare the collaboration propensity of EU27 countries and MIMs among themselves. Other indicators that share a strong correlation are the share of $HCP_{10\%}$ and the SCP, indicating that little is compromised by excluding the SCP from the CEII, as was suggested above. The share of $HCP_{10\%}$ and the SIP are also correlated, but both measure different aspects of science and the correlation is not as strong. Accordingly, both could be considered for inclusion in the CEII.

Table 23: Correlation coefficient between the indicators among the EU27 Member States and MIMs

	Pop.	Pubs.	Pubs./ Pop.	SI	SIP	SPP	SOA	$HCP_{10\%}$	SCP	WEC (world)
Pop.	1.00	0.84	-0.34	0.30	-0.51	-0.23	-0.43	0.02	-0.01	0.59
Pubs.	0.84	1.00	-0.18	0.21	-0.38	-0.05	-0.39	0.16	0.17	0.83
Pubs./Pop.	-0.34	-0.18	1.00	0.17	0.42	0.46	0.22	0.38	0.24	-0.06
SI	0.30	0.21	0.17	1.00	-0.28	-0.34	-0.42	-0.05	-0.18	-0.05
SIP	-0.51	-0.38	0.42	-0.28	1.00	0.39	0.42	0.60	0.45	-0.14
SPP	-0.23	-0.05	0.46	-0.34	0.39	1.00	0.43	0.24	0.32	0.04
SOA	-0.43	-0.39	0.22	-0.42	0.42	0.43	1.00	0.02	-0.02	-0.26
$HCP_{10\%}$	0.02	0.16	0.38	-0.05	0.60	0.24	0.02	1.00	0.68	0.40
SCP	-0.01	0.17	0.24	-0.18	0.45	0.32	-0.02	0.68	1.00	0.47
WEC (world)	0.59	0.83	-0.06	-0.05	-0.14	0.04	-0.26	0.40	0.47	1.00

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Table 24 shows the correlation coefficient between the indicators among the EU27 members only; it also includes the STP and the WEC in the Member States' co-publication network. Many of the conclusions drawn in the last paragraph can be drawn here as well.

In addition, Table 24 shows that the SIP and the STP do indeed share a strong correlation (0.79). Consequently, these two indicators are so related that either one of them would be sufficient to explain more than 50 % of the variance of the other. Having them together in the same framework would make them reinforcing each other, and finally, giving them more importance in the results. On these grounds, and because SIP cannot be calculated for all MIMs, it would make sense to exclude SIP from consideration for the CEII. The SPP and the SCP were also highly correlated, which again is not particularly surprising given that most patent assignees are private firms, which would naturally cite the articles they published in their patents, even though the SPP does not include the (few) articles that were published solely by the private sector. However, the correlation between these two indicators was much weaker when including MIMs. This is perhaps due to the fact that relatively more of the EU27's publications in clean energy included the private sector (15.4 %) compared to those of the MIMs (13.5 %), while a higher share of the MIMs' articles were published solely by the private sector (2.7 %), although that share is only marginally smaller for the EU27 (2.4 %). In any case, the SPP and the SCP should normally converge, and thus, it makes sense to exclude the SCP from the CEII given the results for the MIMs.

Table 24: Correlation coefficient between the indicators among the EU27 Member States

	Pop.	Pubs.	Pubs./ Pop.	SI	SIP	STP	SPP	SOA	$HCP_{10\%}$	SCP	WEC (world)	WEC (EU27)
Pop.	1.00	0.96	-0.26	-0.27	-0.06	-0.39	0.03	-0.13	0.17	0.23	0.89	0.90
Pubs.	0.96	1.00	-0.07	-0.21	0.00	-0.35	0.09	-0.12	0.30	0.31	0.94	0.94
Pubs./Pop.	-0.26	-0.07	1.00	0.37	0.50	0.22	0.27	0.21	0.54	0.33	0.10	0.01
SI	-0.27	-0.21	0.37	1.00	-0.27	-0.30	-0.42	-0.27	-0.11	-0.37	-0.28	-0.34
SIP	-0.06	0.00	0.50	-0.27	1.00	0.74	0.57	0.48	0.64	0.66	0.23	0.21
STP	-0.39	-0.35	0.22	-0.30	0.74	1.00	0.56	0.29	0.34	0.50	-0.22	-0.17
SPP	0.03	0.09	0.27	-0.42	0.57	0.56	1.00	0.43	0.46	0.66	0.26	0.25
SOA	-0.13	-0.12	0.21	-0.27	0.48	0.29	0.43	1.00	0.20	0.40	0.03	0.05
$HCP_{10\%}$	0.17	0.30	0.54	-0.11	0.64	0.34	0.46	0.20	1.00	0.67	0.43	0.44
SCP	0.23	0.31	0.33	-0.37	0.66	0.50	0.66	0.40	0.67	1.00	0.46	0.46
WEC (world)	0.89	0.94	0.10	-0.28	0.23	-0.22	0.26	0.03	0.43	0.46	1.00	0.96
WEC (EU27)	0.90	0.94	0.01	-0.34	0.21	-0.17	0.25	0.05	0.44	0.46	0.96	1.00

Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

In short, the following indicators for the publication dimension could be considered for inclusion in the CEII:

- Number of publications per capita
- Specialisation index
- Share of international co-publications
- Share of open access publications
- Share of public/private co-publications
- Share of publications among the 10 % most cited

If the CEII is to focus exclusively on R&I performance in the traditional sense, this set of candidate indicators could be reduced to the number of publications per capita and the share of publications among the 10 % most cited.

ANNEX 1 SUMMARY OF INDICATORS

Indicator 1: Number of scientific publications

Aspect	Description
Indicator	Number of scientific publications
Description	The number of scientific publications, also referred to as output, is measured for each EU27 member and MIM. It is fractioned across authors and KAs, and normalised by capita.
Rationale/ relevance	The number of scientific publications is a proxy for the level of research intensity among the EU27 and MIMs in the energy sector.
Comparability	The number of scientific publications per capita enables comparison among all EU27 members and MIMs.
Data availability	The number of scientific publications is measured by using the Scopus database. Publication and population data are available for all countries and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way. Note that variation in the completeness of Scopus in the most recent publication year could lead to comparability issues. However, tests have shown that the latest year is complete or near complete for all countries of interest.
Assessment	The number of scientific publications is a simple and an easy-to-understand way to communicate the level of research intensity of countries. However, it is proportional to the size of countries, which makes a comparison between them possible but not particularly insightful without a normalising metric, such as population. The number of publications per capita is recommended for inclusion in the CEII.

Indicator 2: Specialisation index

Aspect	Description
Indicator	Specialisation index
Description	The specialisation index represents the scientific output of a given entity (e.g. a Member State) in a given research area (e.g. nuclear safety) relative to the intensity in a reference entity (e.g. the world) in the same research area. In other words, when an entity is specialised in an area, it puts more emphasis, relative to the reference entity, on that area at the expense of others. Specialisation is therefore said to be a zero-sum game: the more an entity specialises in an area, the less it does in another. Fractional counting of publications across authors and KAs is used to ensure a true zero-sum game. It is calculated as follows: $$SI = \frac{X_S/X_T}{N_S/N_T}$$ Where: X_S Number of publications in for entity *X* in a given research area (e.g. Belgium publications in nuclear safety) X_T Total number of publications for entity *X* (e.g. all Belgium publications) N_S Number of publications for reference entity *N* in a given research area (e.g. world publications in nuclear safety) N_T Total number of publications for reference entity *N* (e.g. all world publications)

Rationale/ relevance	The specialisation index is used as a proxy to identify the countries among the EU27 and MIMs that allocate a significant portion of their resources to the SET Plan KAs, relative to the world. It also helps track those KAs in which they do not allocate resources relatively as much as expected at the world level.
Comparability	The specialisation index enables comparison among all EU27 members and MIMs.
Data availability	The specialisation index is calculated based on the number of scientific publications by using the Scopus database. Data are available for all countries and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The specialisation index is normalised to 1 and is therefore easy to interpret. A score above 1 indicates a level of specialisation above that of the reference entity, and a score below 1 indicates the opposite. The specialisation index is recommended for inclusion in the CEII.

Indicator 3: Share of international scientific co-publications

Aspect	Description
Indicator	Share of international scientific co-publications
Description	The number of international scientific co-publications is the number of publications that include at least two authors affiliated to different countries. The share is the number of international scientific co-publications proportional to the total number of scientific publications. Publications are fractioned across KAs but not authors. It is calculated for each EU27 member and MIM as follows: $$SIP = \frac{X_I}{X_T}$$ Where: X_I Number of publications for entity X (e.g. Belgium) that include at least one author affiliated to another country X_T Total number of publications for entity X
Rationale/ relevance	In the implementation plan of the SET Plan KAs, the EC details at great length the need for Member States to collaborate among themselves, as well as outside the EU, to fulfil the goals of the SET Plan. The share of international scientific co-publications is a proportional measure of that collaboration between countries.
Comparability	The share of international scientific co-publications enables comparison among all EU27 members and MIMs.
Data availability	The share of international scientific co-publications is calculated based on the number of co-publications and the total number of publications by using the Scopus database. Data are available for all countries and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The share of international scientific co-publications is a simple and easy-to-understand way to communicate the proportional level of research collaboration between countries internationally. However, publications are flagged as international co-publications in a binary way if at least two different countries are represented by their authors, no matter how many countries they might represent. Therefore, this indicator does not portray the diversity and relative importance of

	partners. Larger countries also tend to have more opportunities to collaborate internally than smaller countries and, as a result, are more self-sufficient and less likely to collaborate outside their own borders. In this respect, a relatively small share of international co-publications does not necessarily equate to an unwillingness to collaborate. The share of international co-publications is recommended for inclusion in the CEII.

Indicator 4: Share of transnational scientific co-publications

Aspect	Description
Indicator	Share of transnational scientific co-publications
Description	The number of transnational scientific co-publications is the number of publications that include at least two authors affiliated to different EU Member States. The share of transnational scientific co-publications is the number of transnational scientific co-publications proportional to the total number of scientific publications. Publications are fractioned across KAs but not authors. It is calculated for each EU27 member as follows: $$STP = \frac{X_I}{X_T}$$ Where: X_I Number of publications for entity X (e.g. Belgium) that include at least one author affiliated to another Member State X_T Total number of publications for entity X
Rationale/ relevance	In the implementation plan of the SET Plan KAs, the EC details at great length the need for Member States to collaborate among themselves to fulfil the goals of the SET Plan. The share of transnational scientific co-publications is a proportional measure of that collaboration between Member States.
Comparability	The share of transnational scientific co-publications enables comparison among all Member States, but not between MIMs.
Data availability	The share of transnational scientific co-publications is calculated based on the number of co-publications and the number of publications by using the Scopus database. Data are available for Member States and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The share of transnational scientific co-publications is a simple and easy-to-understand way to communicate the proportional level of research collaboration between Member States. However, publications are flagged in a binary way as transnational co-publications if at least two different Member States are represented by their authors, no matter how many Member States they might represent. Therefore, this indicator does not portray the diversity and the extent of the collaborations. Larger countries also tend to have more opportunities to collaborate internally than smaller countries and, as a result, are more self-sufficient and less likely to collaborate outside their own borders. In this respect, a relatively small share of transnational co-publications does not necessarily indicate an unwillingness to collaborate. The share of transnational co-publications is not recommended for inclusion in the CEII, chiefly for the simple reason that it is not calculated for MIMs, but also because it strongly correlates with the share of international co-publications, making it redundant.

Indicator 5: Weighted eigenvector centrality in the world's co-publication network

Aspect	Description
Indicator	Weighted eigenvector centrality in the world's co-publication network
Description	The weighted eigenvector centrality is a network indicator that measures the level of integration of an actor in a collaboration network. It integrates the number of actors to which a given actor is connected (through co-authorship of publications), the intensity of those connections (number of co-publications between two actors), and the importance of the partnering actors to the network structure (connections to hubs are valued more than connections to peripheral actors). Scores for this indicator range from 0 to 1, with 1 representing the most important actor to the network structure (typically a major hub with strong connections to a larger number of actors including other influential players in the network) and 0 representing isolated entities (actors which are disconnected from the network). This indicator is calculated for each of the EU27 members and MIMs using the world's collaboration network (i.e. using all countries). Full counting of co-publications is used.
Rationale/ relevance	In the implementation plan of the SET Plan KAs, the EC details at great length the need for Member States to collaborate among themselves, as well as outside the EU, to fulfil the goals of the SET Plan. The weighted eigenvector centrality is proposed as a possible alternative to the share of international co-publications, which was requested in the ToR. The former indicator captures, by design, a broader range of aspects underpinning international collaboration than the latter. While the latter indicator simply tracks the frequency of publications by a given country that were produced with at least one international partner, the former indicator additionally captures the diversity of involved partners and their relative importance to the network structure in a given research area.
Comparability	The weighted eigenvector centrality enables comparison among all EU27 members and MIMs.
Data availability	The weighted eigenvector centrality is calculated using co-publication data as indexed in the Scopus database. Data are available for all countries and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The weighted eigenvector centrality of actors in a network is benchmarked against the most central actor in the network, no matter how connected or fragmented the network might be. This has to be accounted for in interpreting differences in the centrality of a given country across KAs. A greater score in a given KA compared to another KA could in theory be associated with a smaller share of international co-publications in the former KA if the overall level of international cooperation is also smaller in the former KA. Larger countries also tend to have more opportunities to collaborate internally than smaller countries and, as a result, are more self-sufficient and less likely to collaborate outside their own borders. However, given their sheer size, they tend to be the most frequent collaboration partners of every other country and usually have a high score. For this reason, the weighted eigenvector centrality is not recommended for inclusion in the CEII. The analyses have indeed shown that it correlates strongly with the number of publications and GDP.

Indicator 6: Weighted eigenvector centrality in the Member States' co-publication network

Aspect	Description
Indicator	Weighted eigenvector centrality in the Member States co-publication network
Description	Same as Indicator 5 except that the centrality is computed using a network in which only the 27 Member States are included as actors.
Rationale/ relevance	In the implementation plan of the SET Plan KAs, the EC details at great length the need for Member States to collaborate among themselves to fulfil the goals of the SET Plan. The weighted eigenvector centrality within the co-publication network of the Member State is proposed as a possible alternative to the share of transnational co-publications. The former indicator captures, by design, a broader range of aspects underpinning transnational collaboration than the latter. While the latter indicator simply tracks the frequency of publications by a given Member State that were produced with at least one other Member State, the former indicator additionally captures the diversity of involved Member State and their relative importance to the network structure in a given research area.
Comparability	The weighted eigenvector centrality enables comparison among the EU27 but not the MIMs.
Data availability	The weighted eigenvector centrality is calculated using co-publication data as indexed in the Scopus database. Data are available for all 27 Member States and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The weighted eigenvector centrality of actors in a network is benchmarked against the most central actor in the network, no matter how connected or fragmented the network might be. This has to be accounted for in interpreting differences in the centrality of a given Member State across KAs. A greater score in a given KA compared to another KA could, in theory, be associated with a smaller share of transnational co-publications in the former KA if the overall level of transnational cooperation is also smaller in the former KA. Larger Member States also tend to have more opportunities to collaborate internally than smaller Member States and, as a result, are more self-sufficient and less likely to collaborate outside their own borders. However, given their sheer size, they tend to be the most frequent collaboration partners of every other Member State and usually have a high score. Again, for the same reason given for Indicator 5, but also because it is not calculated for MIMs, the weighted eigenvector centrality in the Member States co-publication network is not recommended for inclusion in the CEII.

Indicator 7: Share of open access scientific publications

Aspect	Description
Indicator	Share of open access scientific publications
Description	The number of open access scientific publications is the number of publications that are publicly and freely available online without any barriers, either through the publisher (known as gold open access), or through a repository or a personal website (known as green open access). The share of open access scientific publications is the number of open access scientific publications proportional to the total number of scientific publications. Publications are fractioned across KAs but not authors. It is calculated for each EU27 member and MIM, as follows: $$SOA = \frac{X_{OA}}{X_T}$$

	Where: X_{OA} Number of publications for entity *X* (e.g. Belgium) that are available in open access X_T Total number of publications for entity *X*
Rationale/ relevance	The SET Plan prides itself on being a platform for sharing knowledge between key actors. The EC also favours transparency and exchange of information to avoid unnecessary duplication of efforts and to stimulate cooperation and coordination in order to achieve the goals of the SET Plan. Open access facilitates the distribution and circulation of knowledge and discoveries, particularly among actors with restricted financial resources. The share of open access scientific publications is a proportional measure of the openness in research.
Comparability	The share of open access scientific publications enables comparison among all EU27 members and MIMs.
Data availability	The share of open access scientific publications is calculated based on the number of publications by using the Scopus and the 1findr databases; the latter provides the open access status of publications. Data are available for all countries. However, there are restrictions for recent years. As stated above, we recommend applying an exception to the reference year for this indicator due to embargo periods on releasing publications in open access.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way. Scopus and 1findr do not perfectly overlap. 1findr is particularly deficient with respect to its coverage of conference papers. Therefore, the open access status is not defined for some publications, which are not used in computing this indicator.
Assessment	The share of open access scientific publications is a simple and easy-to-understand way to communicate the proportional level of openness in research. The share of open access publications is recommended for inclusion in the CEII.

Indicator 8: Share of public/private scientific co-publications

Aspect	Description
Indicator	Share of public/private scientific co-publications
Description	The number of public/private scientific co-publications is the number of publications that include at least one author affiliated to the public sector (academic, government) and one author affiliated to the private sector (firms, corporations). The share of public/private scientific co-publications is the number of public/private scientific co-publications proportional to the total number of scientific publications. Publications are fractioned across KAs but not authors. It is calculated for each EU27 member and MIM as follows: $$SPP = \frac{X_{PP}}{X_T}$$ Where: X_{PP} Number of publications for entity *X* (e.g. Belgium) that include at least one author affiliated to the public sector and one author affiliated to the private sector X_T Total number of publications for entity *X*
Rationale/ relevance	In the implementation plan of the SET Plan KAs, the EC details at great length the need for the public and private sectors to collaborate during all stages of innovation, from funding and conducting research to commercialisation, to fulfil the goals of the SET Plan. The share of public/private scientific co-publications is a proportional measure of that collaboration between both sectors.

Comparability	The share of public/private scientific co-publications enables comparison among all EU27 members and MIMs.
Data availability	The share of public/private scientific co-publications is calculated based on the number of publications by using the Scopus database. Data are available for all countries and for all years relevant to this study.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The share of public/private scientific co-publications is a simple and an easy-to-understand way to communicate the proportional level of research collaboration between both sectors. The share of public/private co-publications is recommended for inclusion in the CEII.

Indicator 9: Share of scientific publications among the 10 % most cited

Aspect	Description
Indicator	Share of scientific publications among the 10 % most cited
Description	The number of citations received by publications is a proxy for measuring contributions to subsequent knowledge generation; however, because citation practices vary between the disciplines of science, a simple count inevitably creates biases. To correct this shortcoming, the number of citations of each publication is normalised by field, publication type and publication year. This measure is known as the relative citation (RC) score. The 10 % most cited articles are determined by their RC scores instead of their raw number of citations. The share of scientific publications among the 10 % most cited is the number of scientific publications among the 10 % most cited proportional to the total number of scientific publications. Fractional counting across authors and KAs is used, and self-citations are excluded. It is calculated for each Member State and country with Mission Innovation membership as follows: $$\textit{Share of pubs. among } 10\% \textit{ most cited} = \frac{X_{10\%}}{X_T}$$ Where: $X_{10\%}$ Number of publications for entity X (e.g. Belgium) that are among the 10 % most cited according to their RC scores X_T Total number of publications for entity X
Rationale/ relevance	Output and collaboration indicators do not inform on the visibility and impact of publications. The share of scientific publications among the 10 % most cited is a publication-to-publication citation indicator and reflects excellence in scientific publishing. The fifth pillar of the Energy Union on research, innovation and competitiveness aims to support the emergence of scientific breakthroughs in low-carbon and clean energy technologies. This indicator is a proxy for identifying such scientific breakthroughs.
Comparability	The share of scientific publications among the 10 % most cited enables comparison among all EU27 members and MIMs.
Data availability	The share of scientific publications among the 10 % most cited is calculated by using the Scopus database. Data are available for all countries relevant to this study. However, there are restrictions for recent years.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.

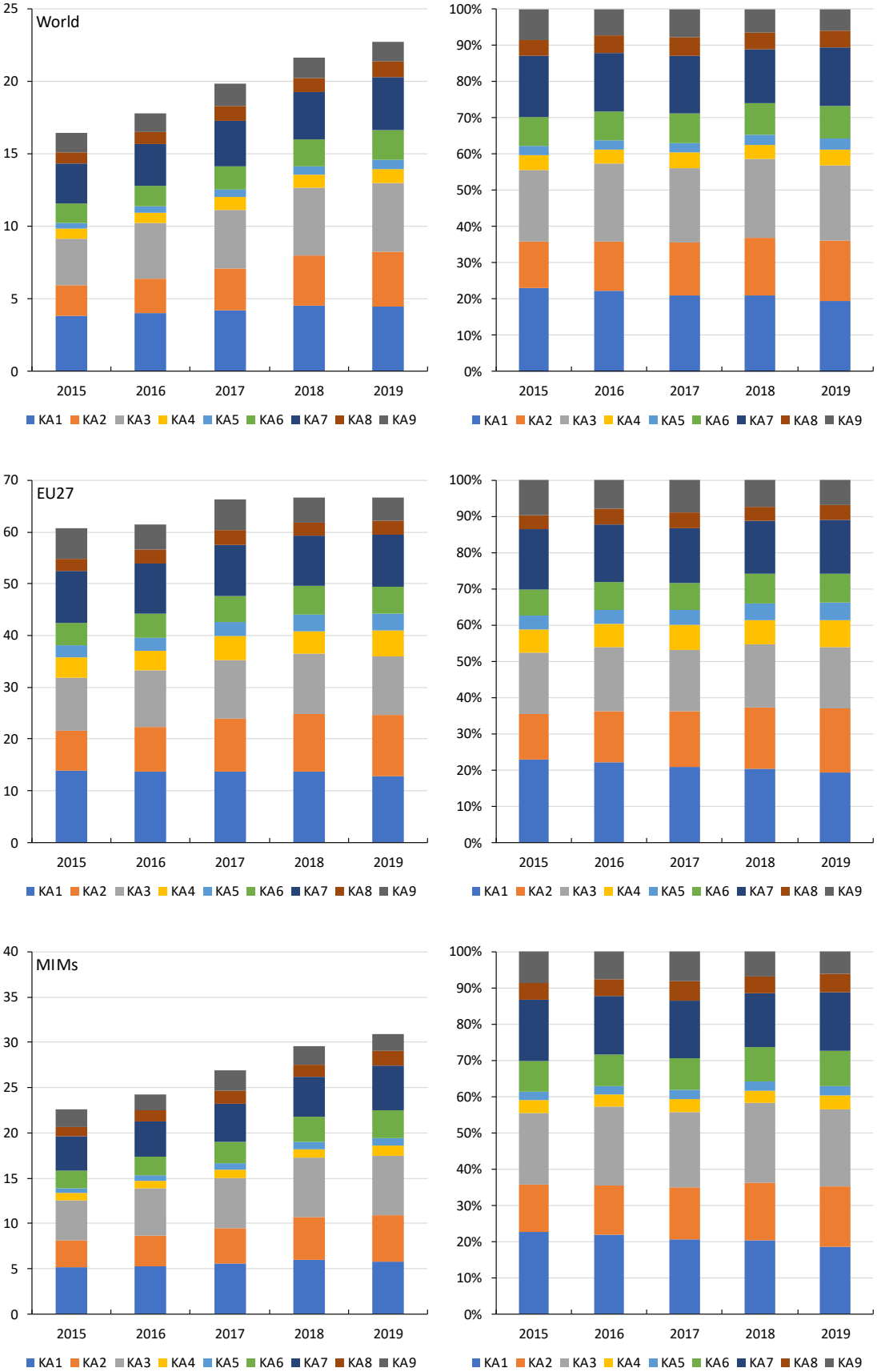
Assessment	The RC score is not calculated for articles published in recent years. Publications need some time to accumulate citations before their impact can be reliably assessed. Typically, three years (the publication year plus the two following years) is granted to publications before their RC score is calculated. Given that conference papers also enjoy far less visibility than journal articles, and are cited far fewer times as a result, the RC score is not calculated for conference papers. A score above 10 % indicates a level of impact above that of the world average, and a score below 10 % indicates the opposite. The share of publications among the 10 % most cited is recommended for inclusion in the CEII.

Indicator 10: Share of scientific publications cited by patents

Aspect	Description
Indicator	Share of scientific publications cited by patents
Description	The number of scientific publications cited by at least one patent is measured for each EU27 member and MIM. The share of scientific publications cited by at least one patent is the number of scientific publications cited by a patent proportional to the total number of scientific publications. Publications are fractioned across KAs but not authors. It is calculated for each EU27 member and MIM as follows: $$SCP = \frac{X_{Pat}}{X_T}$$ Where: X_{Pat} Number of publications for entity *X* (e.g. Belgium) that were cited in at least one patent application X_T Total number of publications for entity *X*
Rationale/ relevance	The share of scientific publications cited in patents is a publication-to-patent citation indicator. Therefore, it crosses the boundaries of scientific publishing by measuring the direct translation of scientific research into innovation. The share of scientific publications cited by patents is a proportional measure of the translation of scientific research into innovation, which is one of the goals of the SET Plan.
Comparability	The share of scientific publications cited in patents enables comparison among all EU27 members and MIMs.
Data availability	The share of scientific publications cited in patents is measured by using and linking the Scopus and the PATSTAT databases, and patents filed at the USPTO, the EPO, the CNIPA, the JPO and the KIPO. Data are available for all countries relevant to this study. However, there are restrictions for recent years.
Quality	Scopus contains some errors with respect to the authors' country affiliation; however, these errors are few and far between and do not compromise the quality of this indicator in any significant way.
Assessment	The share of scientific publications cited in patents is a simple and easy-to-understand way to communicate the relative level of uptake of scientific research into innovation. Citation practices of the technometric literature are also heavily biased toward the technological fields of science. Furthermore, the technometric literature absorbs the scientific literature at a very slow pace compared to the scientific literature itself. A large amount of time is therefore necessary to adequately track the uptake of scientific knowledge in patents. A minimum of five years is usually used. This indicator is therefore not calculated for articles published recently. The share of scientific publications cited in patents is not recommended for inclusion in the CEII for the reason that it correlates well with the share of publications among the 10 % most cited, but specifically because very few publications are ever cited by a patent. This resulted in many EU27 members and MIMs in many KAs with none of their publications cited by a single patent, which makes a comparison between them impractical.

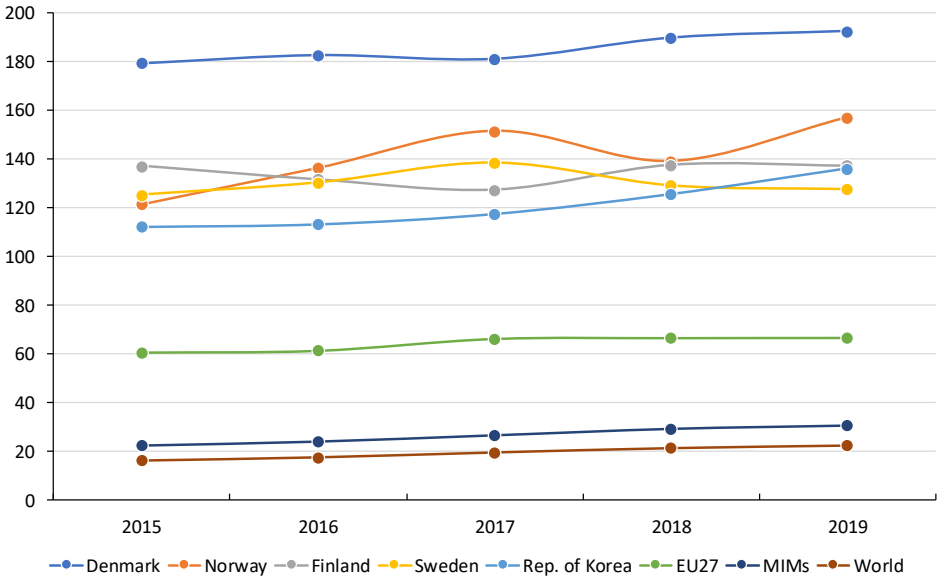
ANNEX 2 ADDITIONAL FIGURES

Figure A2-1: Number of publications per 1,000,000 population by KA (left) and the equivalent proportion (right) for the world, the EU27 and MIMs, 2015-2019



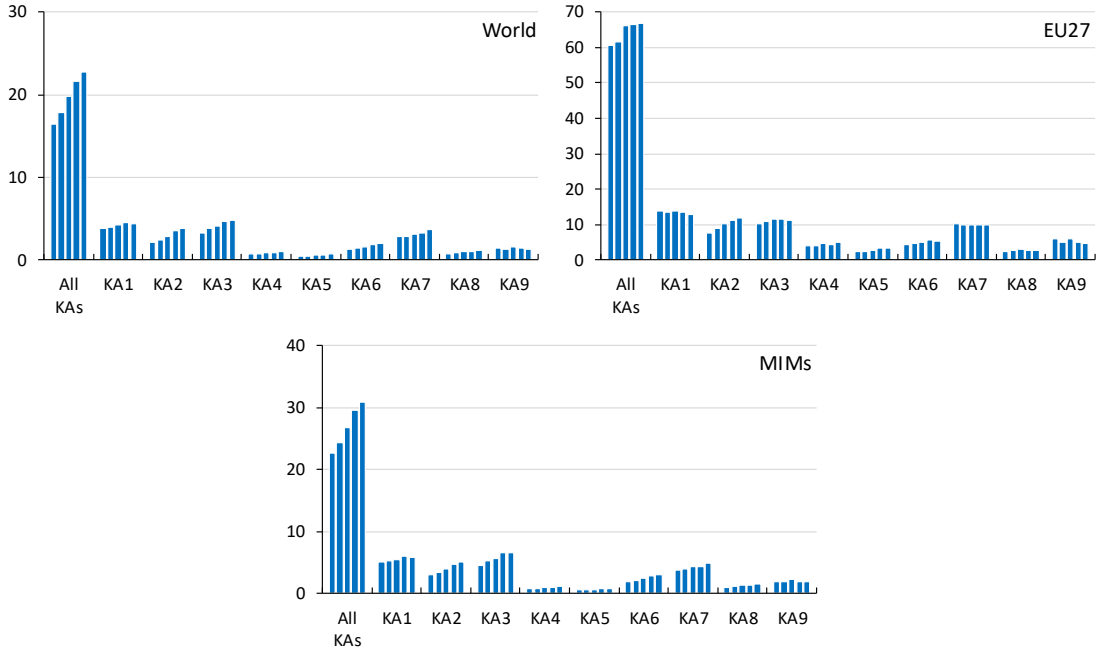
Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Figure A2-2: Number of publications per 1,000,000 population for all KAs combined for the world, the EU27, MIMs and the top 5 countries, 2015-2019



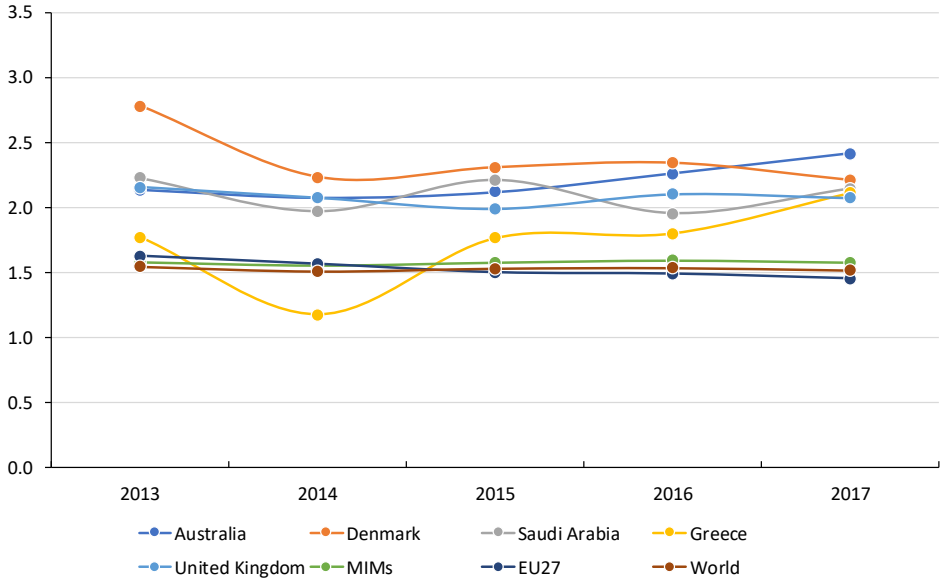
Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Figure A2-3: Yearly number of publications per 1,000,000 population by KA for the world, the EU27 and MIMs between 2015 and 2019



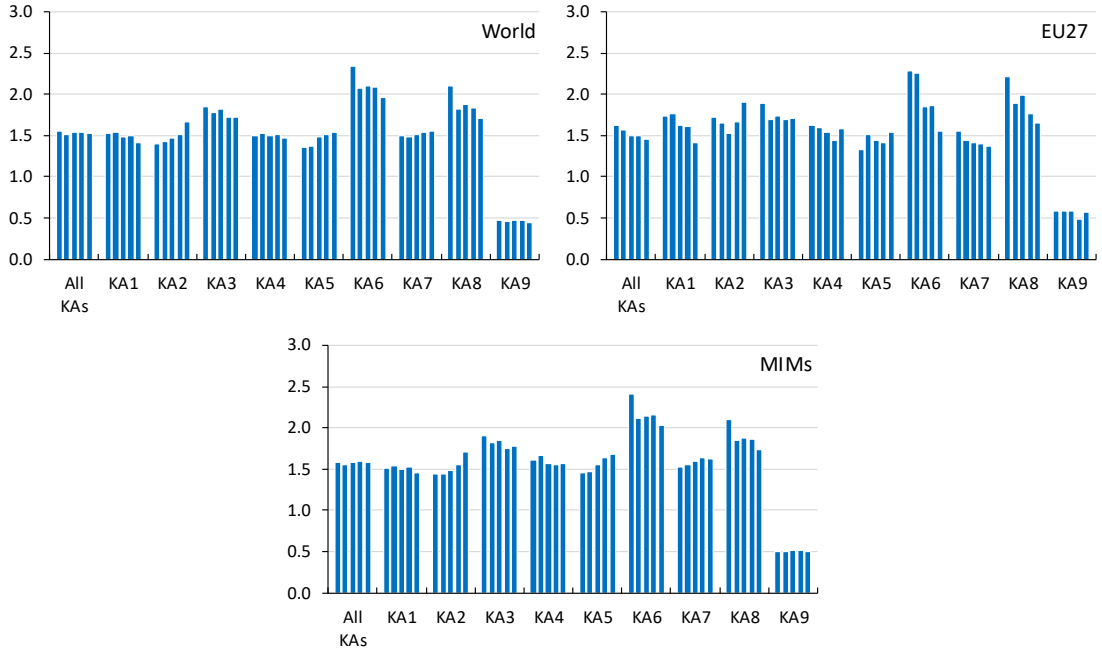
Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Figure A2-4: Share of highly cited publications at the 10 % level for all KAs combined for the world, the EU27, MIMs and the top 5 countries, 2013-2017



Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Figure A2-5: Yearly share of highly cited publications at the 10 % level by KA for the world, the EU27 and MIMs between 2013 and 2017



Source: These statistics were calculated using data from Scopus (Elsevier) and the World Bank

Getting in touch with the EU

IN PERSON
All over the European Union, there are hundreds of Europe Direct information centres.
You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

ON THE PHONE OR BY EMAIL
Europe Direct is a service that answers your questions about the European Union.
You can contact this service:
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696, or
– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

ONLINE
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU PUBLICATIONS
You can download or order free and priced EU publications from:
https://op.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en)

EU LAW AND RELATED DOCUMENTS
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU
The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

The Clean Energy Innovation Index (CEII) is a composite index designed to track progress in achieving the SET Plan key actions, as measured through the lens of scientific publications, patents and trade. This report focuses on the publication aspect of the CEII. Science-Metrix computed and tested 10 publication-based indicators of potential relevance to the policy context surrounding the development of the CEII.

Studies and reports

Publications Office
of the European Union