Results of a bibliometric analysis of research content contributing to the Sustainable Development Goals

Authors and contributors:
Harald Wirsching, VP Strategy and Market Intelligence, Springer Nature
Jos Franssen, Collection & Information Specialist, Maastricht University
Maurice Vanderfeesten, Innovation Manager Research Services, VU Amsterdam
Astrid van Wesenbeeck, Open Science Officer, KB National Library of the Netherlands
Darco Jansen, Manager Open Access & Open Science, VSNU

1. Summary

This analysis is part of a partnership between the Association of Universities in the Netherlands (VSNU), the Dutch University Libraries and the National Library consortium (UKB) and Springer Nature, and has the objective to assess the volume, utilisation and impact that research outputs contributing to the various Sustainable Development Goals have. It includes a comparative element of open access versus non-open access SDG content as well as country comparisons, with a special focus on the Netherlands.

In total, 1.9 million documents that were published between 2015 and 2019 and contribute to the SDGs were analysed using the Dimensions database. Almost 30 thousand of them have at least one author who is affiliated with a Dutch institution. When we compare documents with a Dutch affiliation with the global numbers, we can see the following patterns:

- The countries that published the most SDG-related content between 2015 and 2019 are – not surprisingly – the US and China, the Netherlands ranks 14th.

- The share of SDG-related of all content for the Netherlands is 9%, which is slightly above the global average of 8%, but well below the share of other countries like Australia (12%), Sweden (12%) or the UK (11%).

- Although the number of documents increased by a compound annual growth rate (CAGR) of 12% in the Netherlands between 2015 and 2019, many countries have been growing their SDG-related content faster. Globally the annual growth rate was 15%, in the EU-27 it was 13%.

- Compared to the world average, the Dutch share of publications related to SDG 3: Good Health & Well Being, SDG 13: Climate Action, SDG 11: Sustainable Cities & Communities and SDG 2: Zero Hunger is much higher. On the other hand, the Netherlands contributes relatively little to SDG 7: Affordable & Clean Energy and SDG 4: Quality Education.

- The share of open access content with at least one Dutch author (46%) is much higher than the global average (38%) and the average of the 27 European Union countries (34%). The Dutch leadership role in open access publishing is clearly reflected by these numbers. However, while almost 50% of the Dutch journal content is available as Gold OA, the percentage is only 8% for book chapters and 1% for proceedings papers.

In addition to the volume of SDG-related content from the Netherlands, we also looked at the 'utilisation' of these publications. We found that there is a notable open access advantage for usage and attention of SDG content using both averages and regression models. E.g. the regression model predicted that Gold OA publications in hybrid journals have on average three times as many downloads as their subscription counterparts in the same journals, and an Altmetrics Attention Score that is 1.7 times as high.

For Springer Nature publications, documents from the Netherlands have recorded around twice as many downloads on average than the global figure, and 1.6 times as high as the EU-27 average, surely an effect of the Springer Compact agreement for Hybrid OA journals that has been in place between Dutch universities and Springer Nature since 2015.

2. Introduction and methodology

In 2019, Springer Nature partnered with the Association of Universities in the Netherlands (VSNU) and the Dutch University Libraries and the National Library consortium (UKB) to jointly explore and answer the question "Is open research facilitating progress on global, grand societal challenges?" The UN's SDGs were chosen as some of the world's most pressing challenges, from good health to peace, justice, and strong institutions. In particular, the partnership aimed to provide data, insights and tools to advance knowledge about how academic research and open research is having a societal impact and accelerating the delivery of the United Nations' Sustainable Development Goals (SDGs).[1]

This analysis had the objective to assess the volume, utilisation and impact that research outputs contributing to the various SDGs have. It includes a comparative element of open access versus non-open access SDG content as well as country and institutional comparisons, with a special focus on the Netherlands.

Creating an SDG content corpus
As a first step in this project, the team developed a prototype for mapping scholarly content against five of the UN's SDGs, with Digital Science chosen as a technology partner. Keyword search strings for five SDGs were defined with input from subject matter experts to produce training sets based on publications from the Dimensions platform. The training sets were then used to apply natural language processing and supervised machine learning, resulting in a classification scheme based on five SDGs: SDG 3: Good Health and Well-being; SDG 4: Quality Education; SDG 7: Affordable and Clean Energy; SDG 11: Sustainable Cities and Communities; and SDG 16: Peace, Justice, and Strong Institutions. Initial results from this work was released in December 2019.[2]

In early 2020, Digital Science applied the resulting method and algorithm to the outstanding 12 goals, releasing results for all 17 goals in April 2020, and making these freely and permanently accessible via Dimensions.[3] As a result of this mapping, it was possible to undertake a closer look at this complete corpus of SDG content. SDG-related content from the period 2010 to 2019 was downloaded from the Dimensions interface in May 2020 and further enriched with usage data from Springer Nature and data from Altmetrics. In addition, only content containing information in all necessary meta data fields and defined as one of three publication types - article, proceeding or chapter - were considered.

The mapping of research content to the SDGs has been the subject of several projects, including the SDG bibliometric analysis of the Aurora Universities Network or the SDG dashboard of the VU Amsterdam.[4,5] The results of these projects depend heavily on the methodology chosen as well as the interpretation and translation of the SDGs into relevancy mappings. In general, three different methodologies can be applied: i) content that explicitly mentions the SDGs, ii) a set of keywords

---

[1] For further information, please see the homepage of the project at: www.springernature.com/gp/researchers/sdg-impact

[2] Wastl, Jürgen, & Diwersy, Mario. (2019). Phase 1 and Phase 2 Summary of SDG Project by Springer Nature, VSNU/UKB, Digital Science. Zenodo. http://doi. org/10.5281/zenodo.3904447

[3] Dimensions includes new research category filters for Sustainable Development Goals https://www.dimensions.ai/blog/dimensions-includes-new-research-category-filters-for-sustainable-development-goals/

[4] SDG analysis: bibliometrics relevance https://aurora-network.global/project/sdg-analysis-bibliometrics-relevance/
5 Universiteitsbibliotheek VU ontwikkelt SDG-dashboard voor toetsen maatschappelijke VN-doelen https://ub.vu.nl/nl/nieuws-agenda/nieuwsarchief/2019/okt-dec/ universiteitsbibliotheek-vu-ontwikkelt-sdg-dashboard-voor-toetsen-maatschappelijke-vn-doelen.aspx

that try to 'translate' the SDG targets into search strings (the method currently used by the majority of other SDG relevancy mapping projects), and iii) a supervised machine learning algorithm which again is based on keyword search strings (the method chosen for this project).

In a recent blog post, Ismael Rafols from the Centre for Science and Technology Studies (CWTS) in Leiden highlighted the differences in approaches and concluded that "indicators on the contributions of science to the SDGs are not (yet) robust".[6] Researchers from the University of Bergen came to a similar conclusion when they compared the results of keyword search strings that they developed with the SDG classifier from Elsevier used in their SciVal product (also based on keyword search strings). They found little overlap between the two result sets, although they used the same 'keyword search string' methodology, and concluded that "currently available SDG rankings and tools should be used with caution at their current stage of development."[7]

Despite these limitations, we believe that this project contributes i) to the further development of robust methods to map research content to the SDG goals and ii) to the discussions in the Netherlands about how research funded and performed in the country contributes to the SDGs.

3. SDG-related content from the Netherlands

3.1. Number of documents by country and SDG

The 17 Sustainable Development Goals were adopted by the United Nations General Assembly as part of the 2030 Agenda for Sustainable Development in 2015. Applying the previously described methodology to the Dimensions database results in 1.9 million documents that were published between 2015 and 2019 and contribute to the SDGs. That's 8.1% of all documents in the Dimensions database fulfilling the same criteria.

The number of documents that are assigned to multiple SDGs is very low; only 6.5% of all documents have multiple classifications. Of those, the majority of documents (54%) were assigned to both SDG 7: Affordable and Clean Energy and SDG: 13 Climate Action, all other variations have very little low numbers. This is somewhat surprising considering the general overlap between some SDGs, but could, on the other hand, also be due to the precision of the keyword search strings and the subsequent supervised machine learning algorithm.

The countries that published the most SDG-related content between 2015 and 2019 are – not surprisingly – the US and China. The Netherlands ranks 14th with almost 30 thousand SDG-related documents published between 2015 and 2019.

The share of SDG-related content amongst all publications differs significantly between countries. As mentioned earlier, on average, SDG-related documents made up 8.1% of all documents in the period of 2015-19. Amongst the Top 20 countries with the highest number of SDG-related content, South Africa has the highest share (17.7%), followed by Indonesia (12.6%), Australia (12.2%) and Sweden (11.5%). The three largest producers, USA (9.3%), China (9.0%), and the UK (10.7%) have above average SDG-related content shares, too, as does the Netherlands (9.2%).

---

[6] Consensus and dissensus in 'mappings' of science for Sustainable Development Goals (SDGs)
http://strings.org.uk/consensus-and-dissensus-in-mappings-of-science-for-sustainable-development-goals-sdgs/
7 Armitage,C.S.; Lorenz,M.; & Mikki, S. (2020): Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results? Quantitative Science Studies,1:3, 1092-1108.
https://doi. org/10.1162/qss_a_00071

Contrarily, countries like Germany (6.4%), Japan (5.0%), France (6.1%) or Russia (5.8%) have well below average shares.

Figure 1: Number of SDG-related publications 2015-19 by country of authors – Top 20
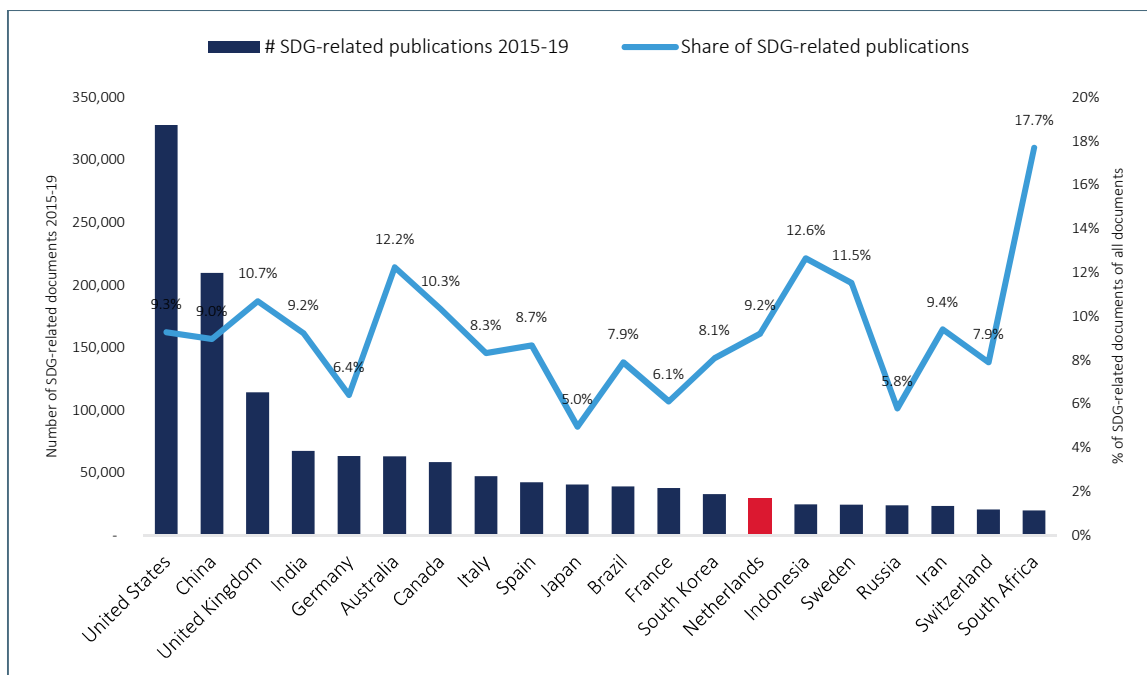
The number of documents published between 2015 and 2019 varies significantly by SDG. Globally, SDG 7: Affordable and Clean Energy (28% of all documents) and SDG 3: Good Health and Well-Being (22%) stand out, comprising together half of all documents. There are three other SDGs that together comprise roughly another third of all documents: SDG 4: Quality Education (11%), SDG 16: Peace, Justice and Strong Institutions (11%) as well as SDG 13: Climate Action (10%). The number of documents in the remaining 12 SDGs is comparably low, altogether they account for only 18% of all documents. Whether this discrepancy is due to the fact that less research is being conducted in these areas or whether the interpretation and translation of the SDGs into relevancy mappings is here more challenging, and therefore less inclusive, is a research question that needs further investigation.

Figure 2: Number of documents published 2015-19 by SDG
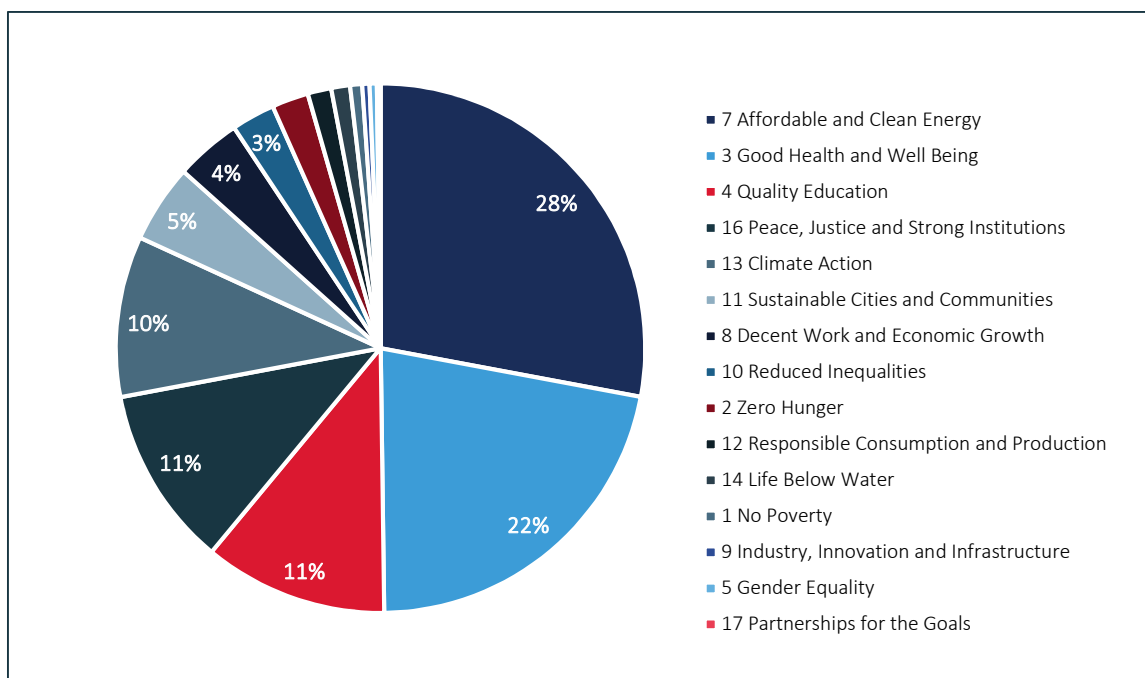
The distribution of content amongst the various SDGs differs substantially between countries, too. Compared to the world average, the Dutch share of publications related to SDG 3: Good Health & Well Being, SDG 13: Climate Action, SDG 11: Sustainable Cities & Communities and SDG 2: Zero Hunger is much higher. On the other hand, the Netherlands contributes relatively little to SDG 7: Affordable & Clean Energy and SDG 4: Quality Education.

Compared to the European Union countries, we see the biggest difference in SDG 7: Affordable & Clean Energy, where the contribution from Dutch researchers is much smaller, while their contribution is much higher in SDG 3: Good Health & Well Being, SDG 16: Peace, Justice & Strong Institutions, SDG 13: Climate Action, and SDG 2: Zero Hunger.

Another interesting observation is that the Dutch research profile with regards to the SDGs is much more similar to the UK than to Germany, Belgium or Denmark.

*Figure 3: Share of SDG-related content 2015-19 by Country/Region*

SDG	World	EU-27	Nether-lands	UK	Germany	Belgium	Denmark
7 Affordable & Clean Energy	27.9%	30.6%	19.4%	21.4%	34.5%	24.2%	32.0%
3 Good Health & Well Being	21.8%	22.0%	28.0%	26.5%	19.8%	30.4%	23.4%
4 Quality Education	11.2%	7.6%	7.4%	7.4%	5.3%	6.0%	5.4%
16 Peace, Justice & Strong Institutions	11.0%	7.2%	10.1%	12.3%	6.3%	8.6%	5.6%
13 Climate Action	9.8%	13.6%	15.0%	12.9%	17.6%	12.4%	17.2%
11 Sustainable Cities & Communities	4.8%	5.4%	5.4%	4.3%	3.9%	4.6%	3.4%
8 Decent Work & Economic Growth	4.0%	3.5%	3.6%	3.5%	3.2%	3.4%	2.7%
10 Reduced Inequalities	2.7%	2.8%	3.1%	4.3%	2.7%	3.0%	2.7%
2 Zero Hunger	2.2%	2.4%	3.2%	2.3%	2.1%	2.8%	2.5%
12 Responsible Consumption & Production	1.4%	1.9%	2.0%	1.6%	1.8%	1.5%	2.3%
14 Life Below Water	1.1%	1.6%	1.2%	1.5%	1.5%	1.2%	1.8%
1 No Poverty	0.7%	0.5%	0.6%	0.8%	0.4%	0.9%	0.4%
9 Industry, Innovation & Infrastructure	0.4%	0.5%	0.4%	0.4%	0.6%	0.3%	0.3%
5 Gender Equality	0.4%	0.4%	0.4%	0.6%	0.3%	0.5%	0.4%
17 Partnerships for the Goals	0.2%	0.1%	0.2%	0.2%	0.1%	0.2%	0.1%

3.2. Number of documents by Publication Type

Globally, 80% of documents indexed in Dimensions, published between 2015-19 and contributing to the SDGs are journal articles, 11% proceedings and 9% book chapters. It could very well be that particularly the book content is underrepresented, since many book publishers do not provide chapter-level data to CrossRef and the major indexing services. In addition, the classification of documents by publication type is not 100% precise. For example, Dimensions doesn't differentiate between different document types in journals (all documents are classified as articles). In addition, proceedings papers published in book formats are classified as chapters and in journal formats as articles.

Nevertheless, as expected, we see significant differences in the various SDGs, which are mainly based on different publishing practices in different disciplines. While in Medicine and Science journal articles are the dominant publication type, proceedings play a much bigger role in Engineering and Computer Science and book publications in the Humanities & Social Sciences. As a result, the SDGs that are closer to Medicine and Science (e.g. SDG 3: Good Health and Well-Being) have the highest share of journal articles, while the ones closest to the Engineering disciplines (e.g. SDG 7: Affordable and Clean Energy) have the highest share of proceedings. And the ones closest to the Humanities & Social Sciences (e.g. SDG 16: Peace, Justice and Strong Institutions) have the highest share of book chapters.

Figure 4: Number of documents 2015-19 by publication type and SDG
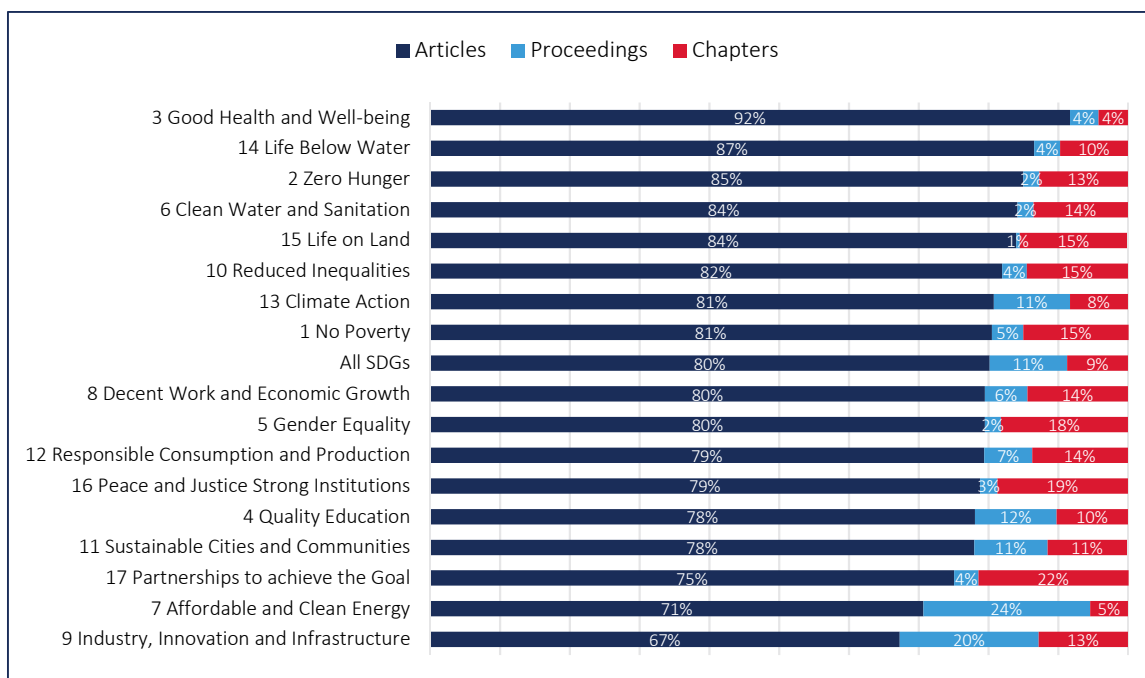
In the Netherlands, the distribution of different document types is even more concentrated in journals than the global average. Journal articles account for 86% of documents, chapters 8% and proceedings the remaining 5%. The low share of proceedings is certainly due to the fact that the Netherlands publishes relatively little in relation to SDG 7: Affordable and Clean Energy, the SDG with the highest numbers of proceedings papers by far.

3.3. Growth in the number of documents

The number of documents contributing to the SDGs has been growing strongly, by a CAGR of 15% between 2015 and 2019. This is much higher than the overall growth of content in the Dimensions database, which has been growing by a CAGR of 7% in that period. Interestingly, the growth rates in the period between 2010 and 2014 have been very similar, a CAGR of 14% of the SDG-related content versus a CAGR of 7% in the whole Dimensions database. So the extraordinary strong growth cannot be attributed to the advent of the SDGs in 2015 alone, although its predecessor, the Millennium Development Goals (MDGs), had similar goals and already started a global effort in 2000.

The number of SDG-related documents with at least one Dutch author increased by a CAGR of 12% between 2015 and 2019. That's below the world average of 15%, but also below the EU-27 average of 13%.

Figure 5: Growth rates of SDG-related content 2015-19 (CAGR)

To put the growth of SDG-related content also in relation to the general growth in output in the various countries, we also looked at the change in percentage of SDG-related content of all content. The share of SDG-related documents in the Netherlands increased from 8.3% in 2015 to 10.6% in 2019, that's an increase of 2.3 percentage points (ppt) in that period. That's slightly below the World and EU-27 averages, but considerably below the increase compared to Denmark, the UK or Belgium.

Figure 6: Share of SDG-related content 2015-19

Country/Year	2015	2016	2017	2018	2019	Ppt change 2015-19
Denmark	9.7%	10.8%	10.9%	12.1%	13.1%	3.4%
UK	9.0%	9.7%	10.5%	11.4%	12.3%	3.3%
Belgium	7.0%	7.9%	8.2%	9.2%	10.1%	3.1%
EU-27	6.8%	7.3%	7.9%	8.6%	9.4%	2.6%
World	6.9%	7.4%	7.8%	8.5%	9.3%	2.4%
Netherlands	8.3%	8.4%	8.9%	9.7%	10.6%	2.3%
Germany	5.6%	6.0%	6.1%	6.8%	7.5%	1.9%

3.4. Number of documents by Open Access status

We clustered all documents with regards to their access status in four different segments:

Figure 7: Classification of access status

Globally, 38% of SDG-related content published between 2015 and 2019 was at the time of this analysis available under a Gold or Bronze Open Access model. The biggest share (17%) was published in a Fully OA publication under an open license. Another 12% of documents are freely available on a publisher platform, but without an open licence (=Bronze OA). Nine percent of documents are published under an open licence in a hybrid publication.

The share of open access content with at least one Dutch author (46%) is much higher than the global average (38%) and the average of the 27 European Union countries (34%). The Dutch leadership role in open access publishing is clearly reflected by these numbers.

*Figure 8: Share of Open Access content by Open Access status 2015-19*
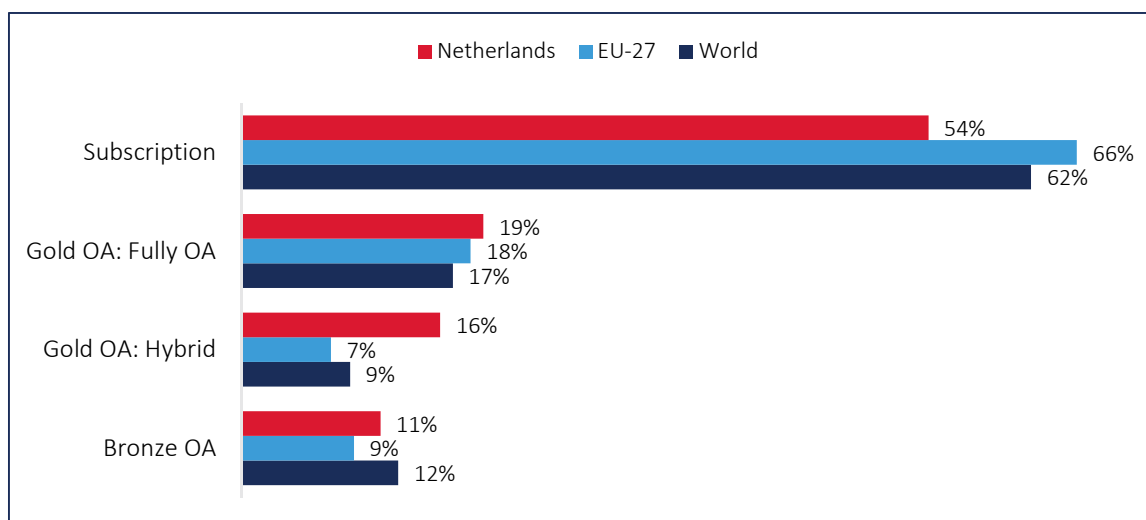
When we look at the share of SDG-related documents that were published in Fully Gold OA publications between 2015 and 2019, we see relatively few differences between the countries selected in our comparison. Overall we see a trend of circa 15% of documents published in Fully Gold OA publication in 2015 that rose to circa 20% in 2019.

Figure 9: Share of SDG-related content published in Fully Gold OA publications
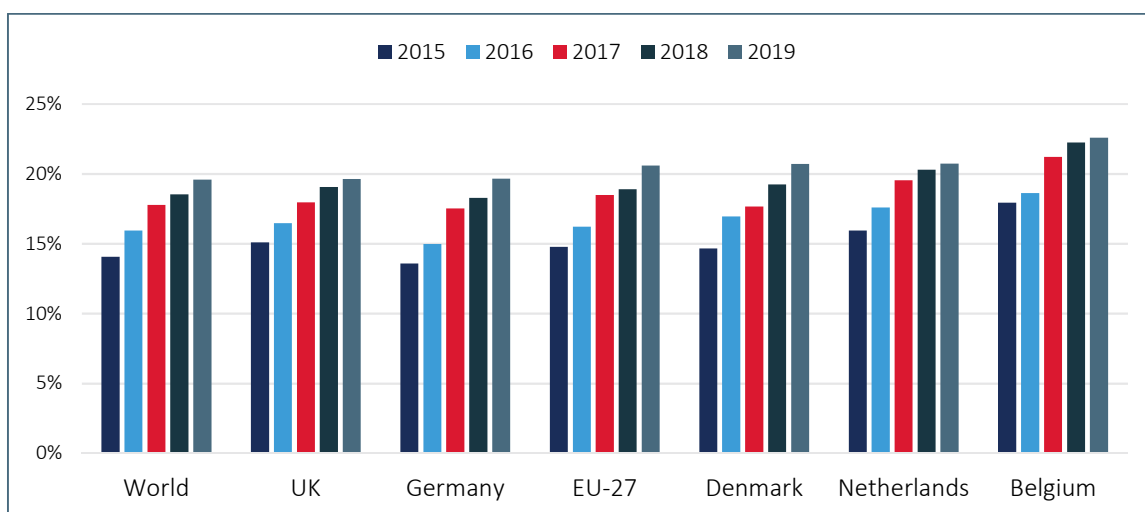

A much bigger difference can be seen when we look at the share of Gold OA documents published in hybrid publications. This number depends to a large degree on funder mandates and policies in the various countries, as well as whether there are 'transformative agreements' in place. Here, the Netherlands stands out. In 2019, more than 20% of all SDG-related documents were published open access in hybrid publications. This is approx. four times as high as the world average, and 2.4 times as high as the EU-27 average. This is certainly due to the many agreements between Dutch university libraries and publishers with an open access element starting in 2015. Currently, there are approx. 11,000 journals in which corresponding authors of Dutch universities and academic hospitals can publish open access for free or with a substantial discount.

Figure 10: Share of SDG-related content published Gold OA in hybrid publications
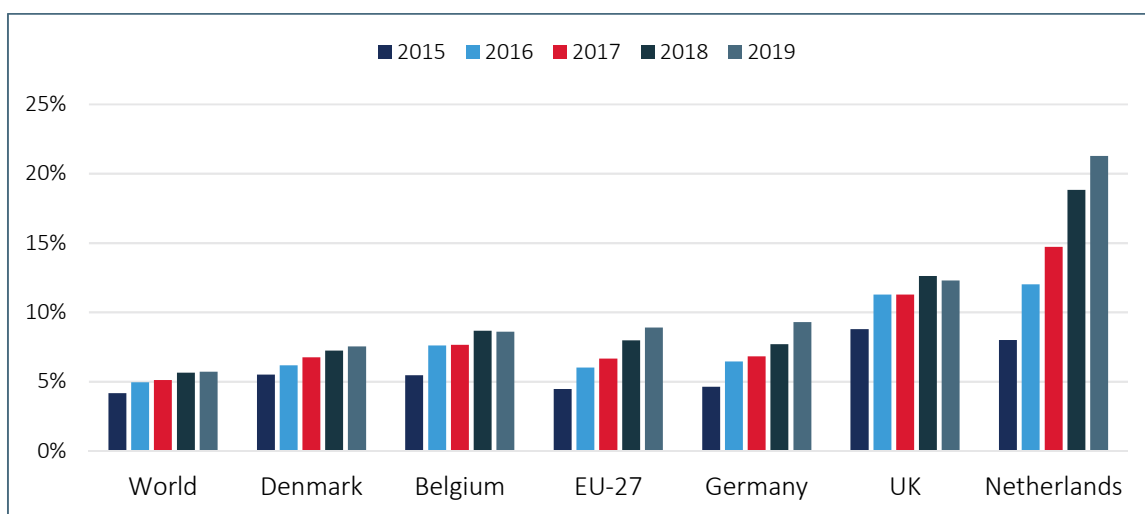
So we've seen that in 2019 a high proportion of SDG-related content with at least one Dutch author was published in Gold OA in either Fully OA or hybrid publications. However, this varies substantially between publication type. While almost 50% of the journal content is available as Gold OA immediately, the percentage is only 8% for chapters and 1% for proceedings. As mentioned previously in this report, the number of chapters and proceedings are very likely significantly underrepresented in the Dimensions database, so there is certainly a significant Dutch research content base that is currently not covered by agreements between Dutch university libraries and publishers.

Figure 11: Share of Dutch SDG-related content available immediately via Gold OA by publication type
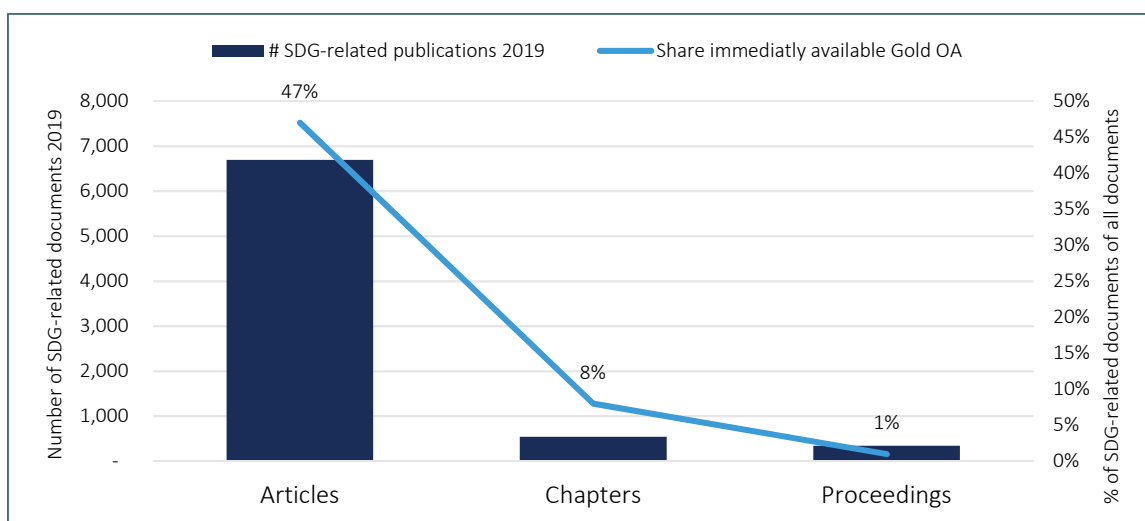
3.5. Utilisation of SDG-related content with a special focus on OA vs. non-OA content

This chapter looks at the utilisation of SDG content in terms of citations, altmetrics attention scores as well as online usage[8]. The latter is limited to Springer Nature content only. Since all metrics build over time, the analysis focuses on publications from a single year in order to guarantee a 'like-for-like' comparison. The publication year 2017 was chosen (as an average time-frame of three years since the metrics were pulled in mid-2020), which seems to be a good compromise in terms of recency on the one hand and sufficient time-frame to build the various metrics on the other.

So when we now look at documents only that were published in 2017, had at least one author who is affiliated with a Dutch institution and are related to at least one SDG, we have a sample of 5,663 total documents and a sample of 1,061 documents published by Springer Nature.

Since the OA status is not the only factor that influences these metrics, we created a model that aims to correct the influence of variables at the document level (SDG, subject field, publication type and whether the document acknowledged any external funding), at the author level (institutional reputation, based on the proxy of the Times Higher Education World University Rankings, and country) and the journal level (Journal Impact Factor, as a proxy for perceived journal prestige). We used negative binomial regression models for all the analyses.

---

[8] Citations numbers are from Dimensions, altmetrics data from Altmetric, and online usage data refers to COUNTER 4 total full-text requests

Online usage:

We see a clear Gold OA usage advantage. For example, the regression model predicts that Gold Hybrid OA documents have on average 3.0 times as many downloads as their Subscription counterparts in the same journals, meanwhile content in Fully Gold OA publications have 1.7 times more.

Altmetrics attention score:

Gold OA content is also shared more often and gets more attention than Subscription content. The average Altmetric Attention Score, as well as the average number of Twitter accounts, news outlets and policy documents mentioning SDG-related content is higher for Gold OA than Subscription content, with an average Altmetric Attention Score that is 1.9 times as high as for Gold Hybrid OA content, and 1.5 for Fully OA content.

Citations:

The data is less distinct when it comes to citations. Although we see higher average citations for Gold OA compared with Subscription content, the results are less clear when controlled for other factors. For example, the model predicts the number of citations to be roughly the same for content in Fully Gold OA publications and Subscription content. However, the model suggests a citation advantage of 1.5 times for documents that are published Gold OA in Hybrid publications.

Figure 12: Downloads, attention and citations of SDG related documents with at least one Dutch author, published in 2017, based on average and regression model predictions

OA Type	Downloads		Altmetrics Attention Score		Citations	
	Average	Regression model predictions	Average	Regression model predictions	Average	Regression model predictions
Subscription	100%	100%	100%	100%	100%	100%
Gold OA: Hybrid	172%	298%	190%	168%	183%	148%
Gold OA: Fully OA	101%	174%	150%	148%	106%	96%

The variation noted between Hybrid OA and Fully OA is worthy of note. Several factors could potentially be at play, including the fact that hybrid journals typically are more established and therefore attract more users. Although we used the Journal Impact Factor as a proxy for journal prestige in our regression model to control for journal reputation, there may be limitations with this approach. In addition, there may be a level of selection bias with the possibility that authors choose OA in a hybrid journal for their most significant work, which therefore attracts higher attention.

When we compare the download numbers from documents with at least one Dutch author with the global and EU-27 ones, we can see that the usage is much higher. Both the average and median number of downloads of Dutch content is more than twice as high as the global average, and 1.6 times as high as the EU-27 average. A main reason for this is the fact that the share of Gold OA content amongst Dutch content in Springer Nature publications is substantially higher than for the global and EU-27 averages, as previously shown, and the open access usage advantage.

Figure 13: Average and median downloads of SDG related Springer Nature documents published in 2017
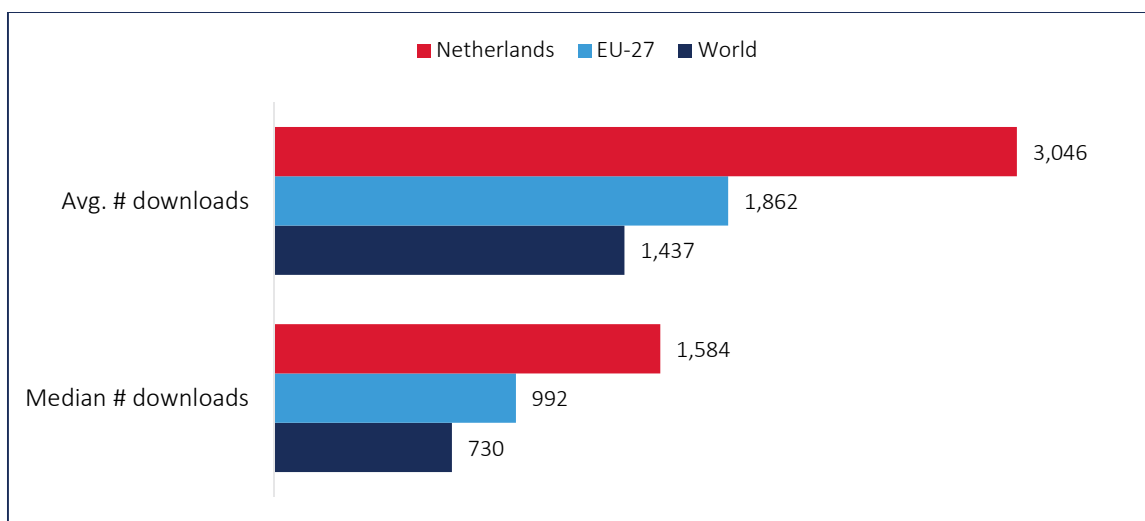
This 'utilisation' analysis has shown that open access content has substantial higher online usage and attention than content that is only available under a subscription model. There are very many studies available that have looked at whether there is an OA citation advantage or not, with mixed results.[9] It is important to note that many of these studies haven't controlled for confounding factors, were limited to just a few journals or a single discipline, and that many found a correlation but didn't suggest a causation. Although this study did find an OA citation advantage for SDG-related content when controlled for selected variables at the document, author and journal level, the effects are smaller when compared to the online usage and attention advantage we observed.

Since citations are mainly an indicator for academic utilisation, this data supports the assumption that OA reaches a substantial number of user groups outside of academia that typically don't have access to a large amount of Subscription content, and are typically less likely to cite this work. Phil Davis concluded in one of his studies that "the real beneficiaries of Open Access may not be the scientific author community … but communities of practice that consume, but rarely contribute to, the corpus of literature. These individuals may include students, educators, physicians, patients, and researchers employed by private industry who depend on the publication of scientific literature."[10]

---

[9] Tennant, J. (n.d.): The Open Access Citation Advantage. Science Open collection. Retrieved 27 Oct 2020 from https://www.scienceopen.com/collection/996823e0-8104-4490-b26a-f2f733f810fb

[10] Davis, P.M. (2010:) Does OA lead to increased readership and citations? The Physiologist, 53(6): 197–201