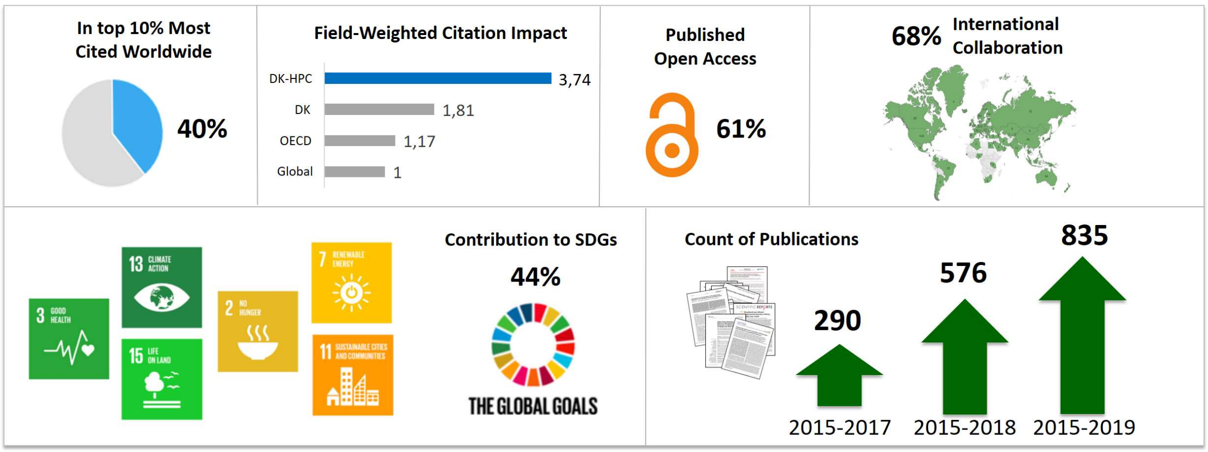
Bibliometric Analysis of Research Output from ABACUS2.0 and Computerome in Denmark, 2015 to 2019

By Dorte Drongstrup and Birgitte Vedel Thage

University Library of Southern Denmark, dh@bib.sdu.dk; DeiC, birgitte.vedel.thage@deiC.dk



Introduction

National supercomputing is a tool widely used in Danish research. Research projects using the national HPC facilities have now published more than 800 international publications.

The Danish national supercomputers were established back in 2014 and 2015, for common benefit, and for Danish research to carry out world-class research moving forward. DeiC is closely following the development of users and research that are using the national supercomputing systems, also called High Performance Computing (HPC). Three national HPC facilities were dedicated to Life Science (Computerome), multidisciplinary research (ABACUS2.02.0) as well as social sciences and the humanities (Cultural Heritage Cluster, KAC), respectively. In this report, we will refer to either ABACUS2.0 or Computerome, when describing the publications using one of the HPC facilities. When referring to both “super computers” the term HPC is used. There is currently only registered two publications from KAC, which may due to the fact that this HPC was fully operational from 2019.

Since the establishment of the national HPC facilities, the number of scientific publications has only increased. This analysis was first published in DeiC's 2017 Annual Report, with a total of 290 publications being recorded for the period 2015-2017. Later, 60 publications were post-registered for 2016-2017. From December 2018 to September 2019, information about 197 publications were registered. In the period January-June 2020 information about 287 publications from 2019 were registered. Thus, a total of 834 publications published in the period 2015-2019 were used as a basis for the analysis.

Method

All the collected data was cleaned and prepared for analysis. The most important identifier used to collect bibliometric information was the DOI (Digital Object Identifier), which is a unique identifier assign to most journal articles since the introduction in 2000. IF the DOI was not available, we used ISBN or Pubmed ID (PMID). Bibliometric information about these publications were collected in the citation databases Scopus and Web of Science.

A search query using DOI, PMID or ISBN was performed in both databases. Of the 834 publications Scopus has indexed 805 publications and Web of Science has indexed 779 publications. Figure 1 shows a comparison of the citation databases, where 92.7% of the publications are indexed in both databases. 3.8% of the publications are only indexed in Scopus, 0.7% are only index in Web Of Science and 2.8% of the publications are indexed in neither databases.

Based on these results, Scopus was selected for data collection. We exported the results of the Scopus query using a CSV file format, as well as exported the results to SciVal for gathering more bibliometric information. Furthermore, we gathered information from the Danish Open Access Indicator website ("the Danish Open Access Indicator," 2020) about Open Access.

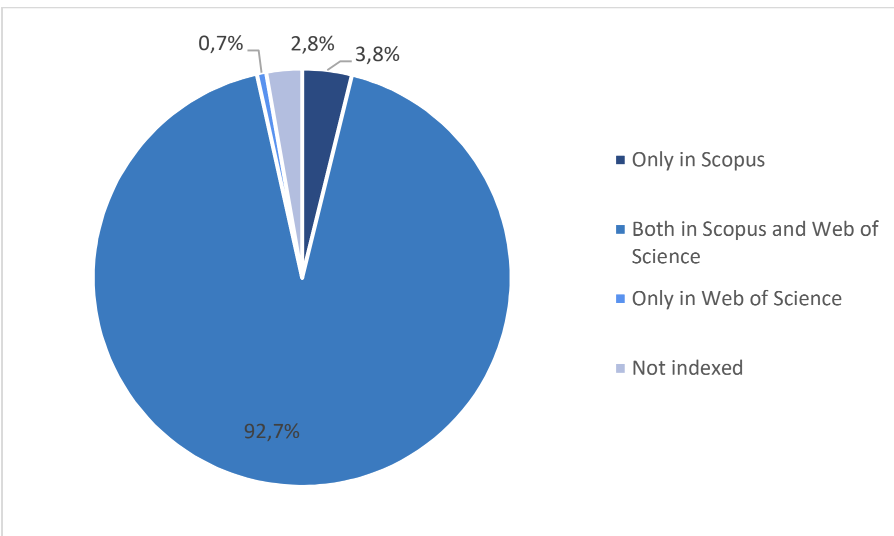


Figure 1 Scopus and Web of Science citation databases coverage

In this report, we use the metric Field-Weighted Citation Impact (FWCI). This metric indicates how the number of citations received by an entity's publications compares with the average number of citations received by all other similar publications in the data universe. Thus, if the FWCI is 1, then the metric indicates that the entity's publications have been cited exactly as would be expected based on the global average for similar publications.

If the FWCI is more than 1, then this indicates that the entity's publications have been cited more than would be expected based on the global average for similar publications. If the FWCI is less than 1, this indicates that the entity's publications have been cited less than would be expected based on the global average for similar publications. Similar publications are defined as those publications in the Scopus database, which have the same publication year, publication type, and discipline. Disciplines are represented by the Scopus classification system (ASJC: All Science Journal Classification). A publication can be assigned to more than one discipline, and then it will be fractionalized based on the number of disciplines assigned.

ANALYSIS

Table 1 displays the distribution of publications from the research projects using ABACUS2.0 and Computerome.

Computerome has more users and a larger output of publications in the period than ABACUS2.0. This can partly be because of the disciplinary differences, which are apparent when comparing Computerome and ABACUS2.0 subject areas in figure 12 and 13.

The majority of ABACUS2.0 publications are from the technical and natural sciences, which has a lower publication and citation rate than the health and life sciences publications, which are 41% of Computerome's publications. Thus, Computerome publications are in average more cited.

Table 1 Summary 2015-2020

SUMMARY	PUBLICATIONS	UNIQUE AUTHORS	CITATIONS	AVERAGE CITATION PER PUBLICATION	FWCI
HPC	805	7,022	24,302	30.2	3.74
ABACUS2.0	278	843	3,182	11.4	1.63
COMPUTEROME	527	6,215	21,120	40.1	4.86

Publication distribution in the different Journal Rankings

We examined the distribution of publications in three journal rankings Citescore Percentile, SCImago Journal Rank (SJR) and Source-Normalized Impact per Paper (SNIP). Figure 2 display the share of HPC publications published in journals, which are included in a journal ranking. Over 96% of the HPC publications are published journals included in all three journal rankings. Each journal ranking indicator is described in the next section.

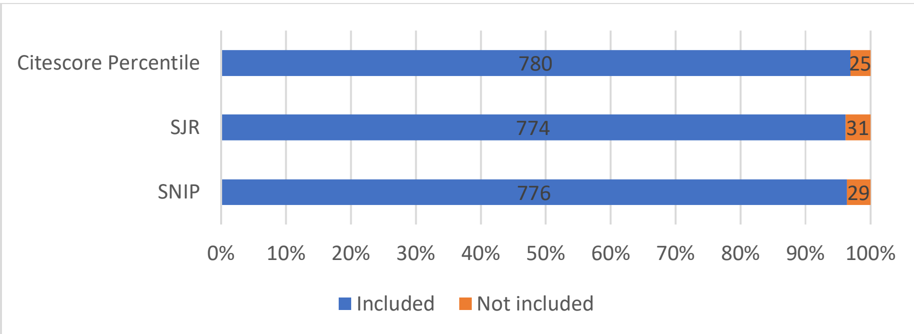


Figure 2 The share of HPC publications publish in journals included in the three journal rankings

Citescore Percentile measures the relative position within subject field based on CiteScore. Citescore is calculated based on the number of citations to documents (articles, reviews, conference papers, book chapters, and data papers) by a journal over four years, divided by the number of the same document types indexed in Scopus and published in those same four years. For more information, see https://service-elsevier-com.proxy1-bib.sdu.dk/app/answers/detail/a_id/14880/supporthub/scopus/.

SJR measures the prestige of citations received by a journal. It weights the value of a citation depending on the field, quality and reputation of the journal that the citation comes from, so that "all citations are not equal." SJR is a numeric value indicating the average number of weighted citations received during a selected year per document published in that journal during the previous three years. The average SJR value for all journals in Scopus is 1.000. Higher SJR values are meant to indicate greater journal prestige. SJR takes differences in the behavior of academics in different disciplines into account and can be used to compare journals in different fields. For more information, see https://www.scimagojr.com/SCImagoJournalRank.pdf.

SNIP measures the citation impact of a journal. SNIP is a ratio between the "Raw Impact per Paper", a type of Citations per Publication calculation, actually received by the journal, compared to the "Citation Potential", or expected Citations per Publication, of that journal's field. SNIP takes differences in disciplinary characteristics into account and can be used to compare journals in different fields. The average SNIP value for all journals in Scopus is 1.000. For more information, see Waltman, van Eck, van Leeuwen, & Visser (2013).

Figure 3, 4 and 5 display the distribution of HPC publications in the top1%-top100%. The figures display how there can be differences in the ranking of journals depending on the indicator. Still, all three figures show how the majority of HPC publications are published in the top25% of Scopus Journal Sources based on all three journal rankings.

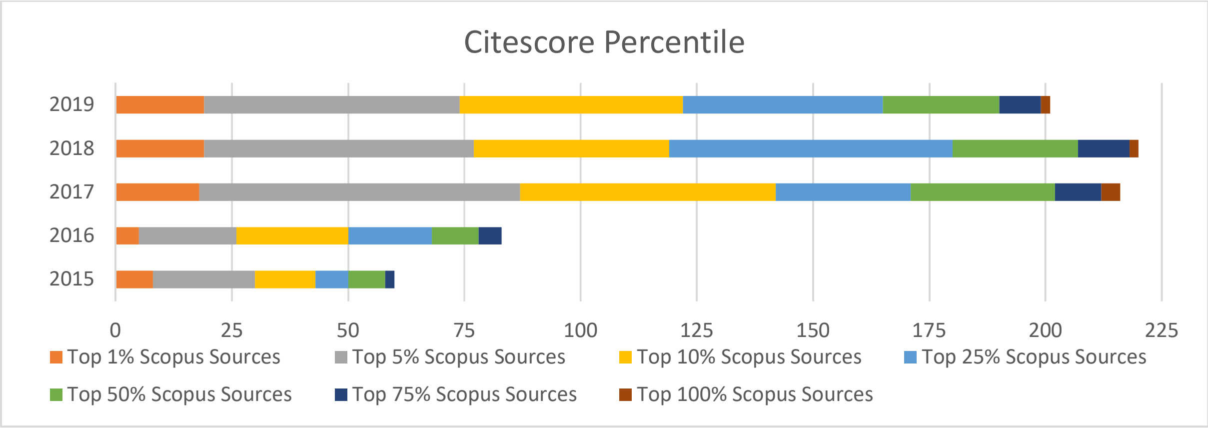


Figure 3 the distribution of HPC publications in the top1%-top100% based on the Citescore Percentile ranking

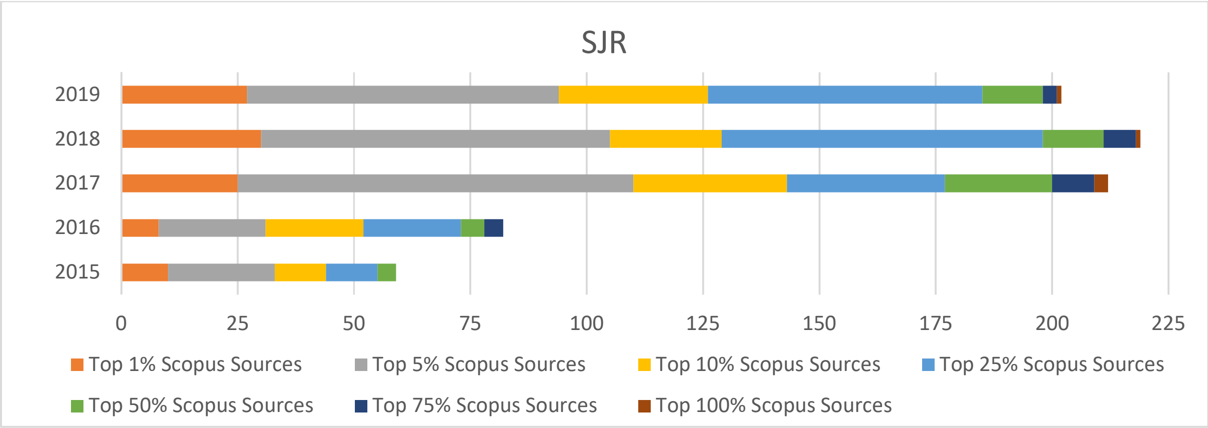


Figure 4 the distribution of HPC publications in the top1%-top100% based on the SJR ranking

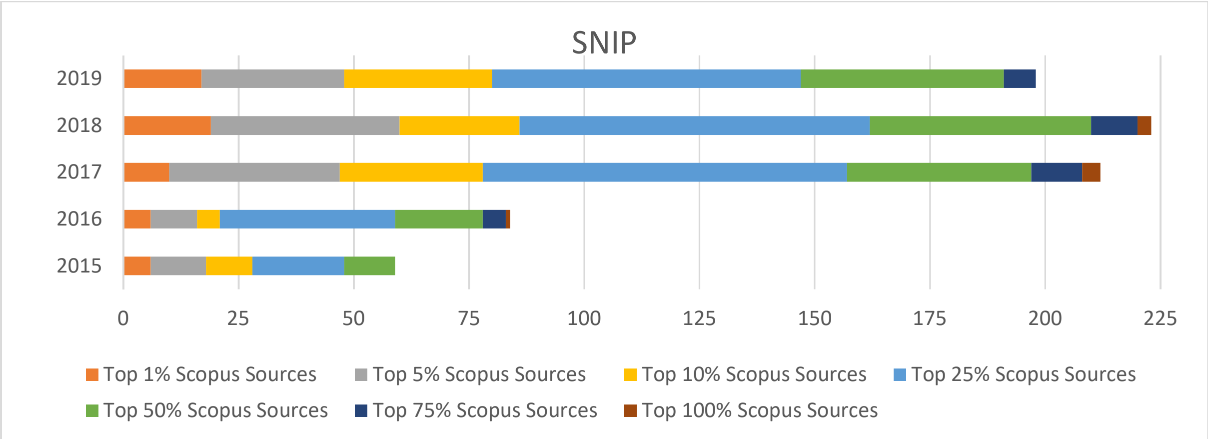


Figure 5 the distribution of HPC publications in the top1%-top100% based on the SNIP ranking

Share of publications in top10% journal ranking

Figure 6 shows the share of danish publications in top10% in all three journal rankings. Danish publications have a tendency to be published in high ranking journals. In all three journal rankings, the share of publications in the top10% are over 20%.

Figure 7 shows the share of HPC publications in top10% in all three journal rankings. Thus, the 59% of the HPC publications are in the top10% in Citescore Percentile ranking. 61% of the HPC publications are in the top10% in SJR ranking, while it is only 36% in the top10% in the SNIP ranking. The share of HPC publications in the top10% ranking journals are higher than the general trend for Danish research publications.

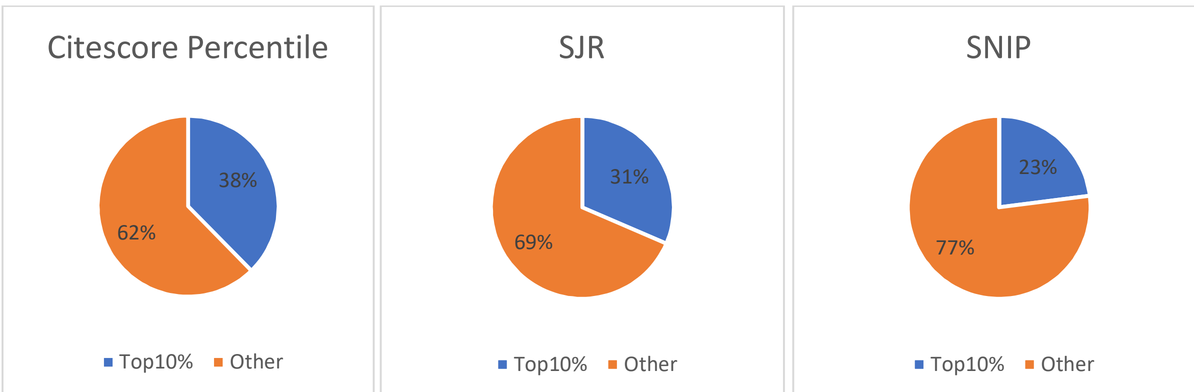


Figure 6 Share of Danish publications in Top10%.

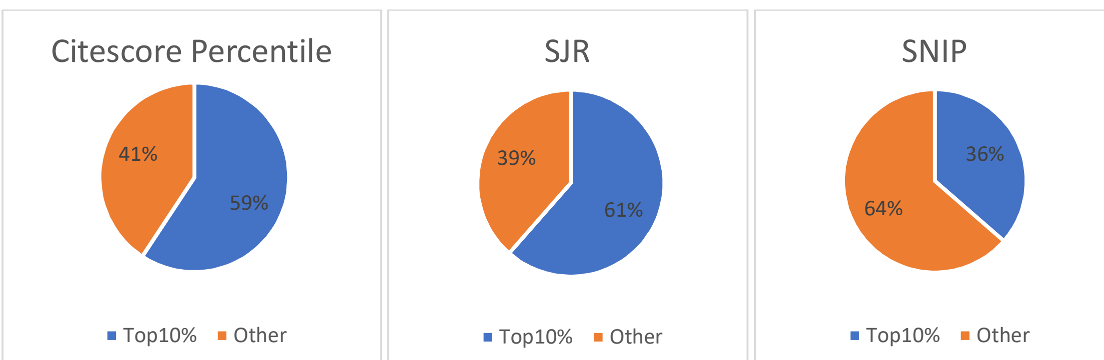


Figure 7 Share of HPC publications in Top10%

Share of publications in Top Citation Percentiles

Figure 8 displays the share of Danish publications that are among the top10% most cited publications worldwide. The "Outputs in Top Citation Percentiles (field-weighted)" is 3% lower than "Outputs in Top Citation Percentiles", which could be because of large share of Danish research is in high-citing areas, such as health and life sciences.

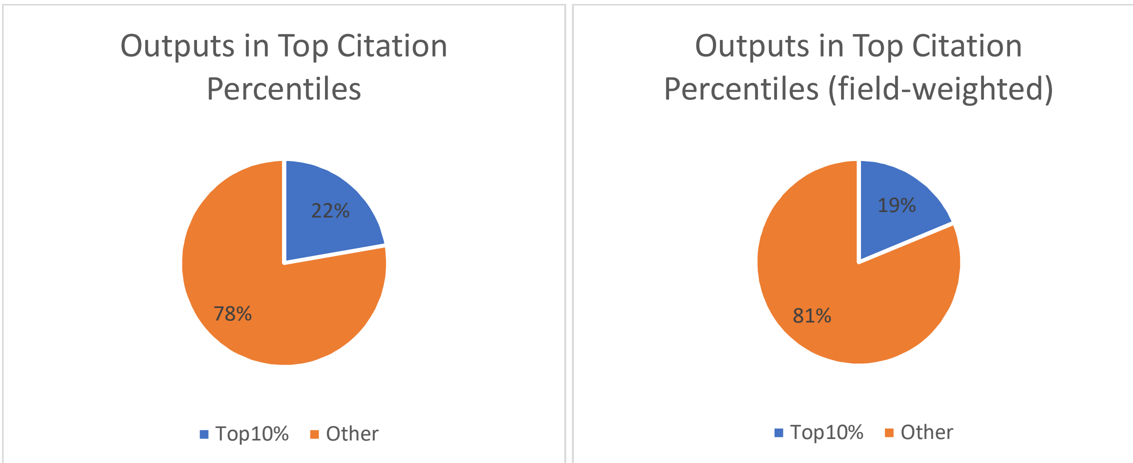


Figure 8 Share of Danish publications in the Top10% Citation Percentile.

Figure 9 illustrate the share of HPC publications among the top10% most cited publications worldwide. The "Outputs in Top Citation Percentiles (field-weighted)" is 13% lower than "Outputs in Top Citation Percentiles". This indicates that most of the research using HPC is from high-citing areas. This is considered by using the field-weighted citation outputs. Nevertheless, the share of HPC publications among the top10% is still substantial and represent 27% of the HPC publications.

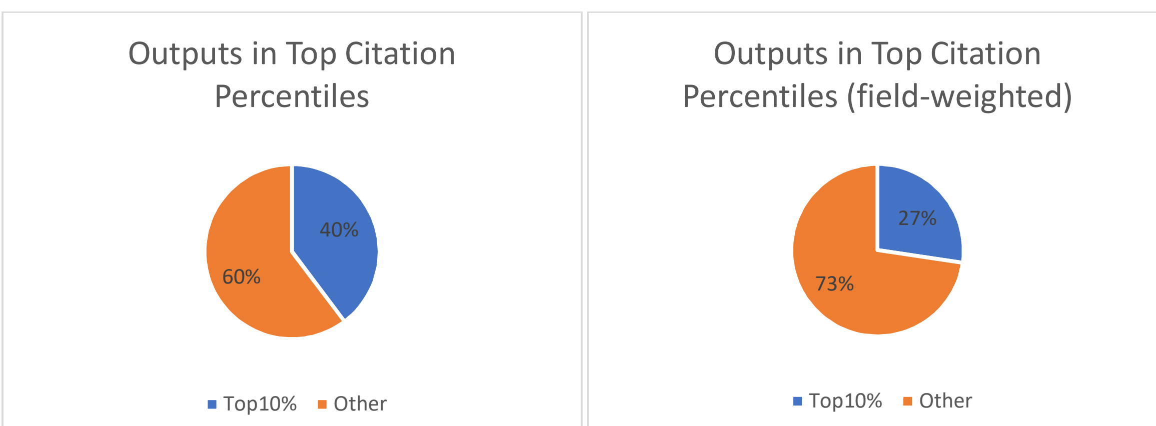


Figure 9 Share of HPC publications in the Top10% Citation Percentile.

Research subject areas

Figure 10 displays the subject area distribution of the HPC publications based on the FORD – Field of Science and Technology (FORD) Classification. FORD is used by OECD to classify research, and have six overall categories: Natural Sciences, Engineering and Technologies, Agricultural Sciences, Medical Sciences, Social Sciences and Humanities. The 805 publications belong mainly to the Natural Sciences and Medical Sciences. 256 of these publications are classified as belonging to more than one subject area (see table 2).

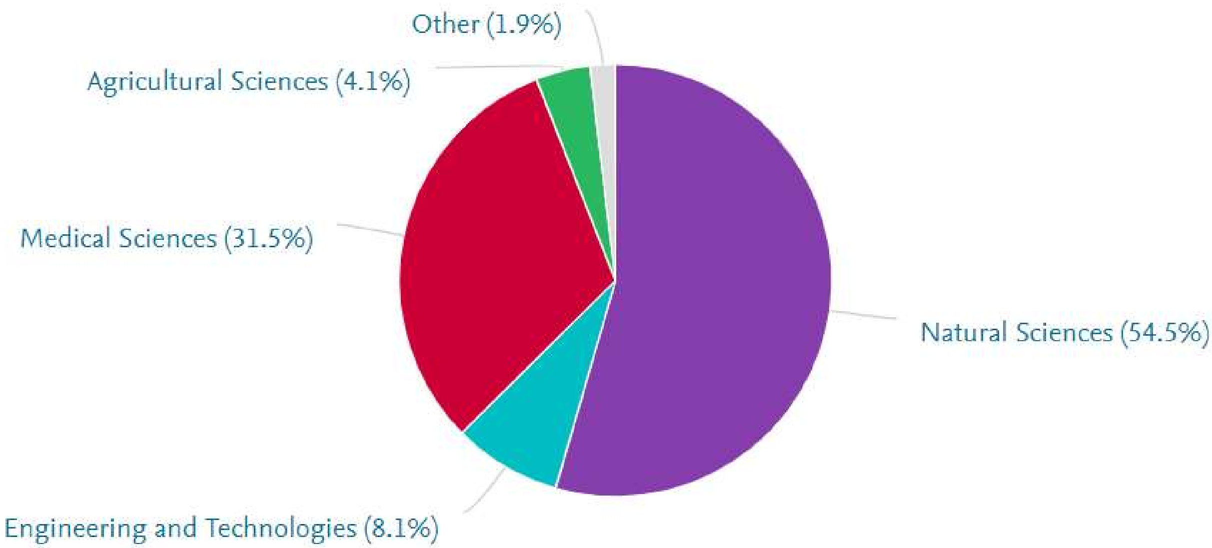


Figure 10 HPC FORD subject areas

Table 2 The FORD subject area distribution of publications

FORD	ABACUS2.0	COMPUTEROME	TOTAL
AGRICULTURAL SCIENCES	6	37	43
ENGINEERING AND TECHNOLOGIES	61	25	86
HUMANITIES	0	1	1
MEDICAL SCIENCES	51	283	334
NATURAL SCIENCES	232	346	578
SOCIAL SCIENCES	14	5	19
TOTAL	**364**	**697**	**1061**
TOTAL PUBLICATIONS WITHOUT DUBLICATES	**278**	**527**	**805**

Figure 11 displays the subject area distribution of the HPC publications based on WoS Categories content, but manually assigned by DeiC into "Scientific Fields". These eight categories are used on DeiC's website to categorize publications (see also https://vidensportal.DeiC.dk/en/HPC/literature/search). If we compare figure 10 and 11, it is evident that the medical sciences are still dominant in both classifications, while FOS's Natural Science category are divided into more categories in DEIC's Scientific Fields (Physics, Biology, Chemistry and Biochemistry, Computer Science & AI). Still, in this report we are primarily using the FORD Classification.

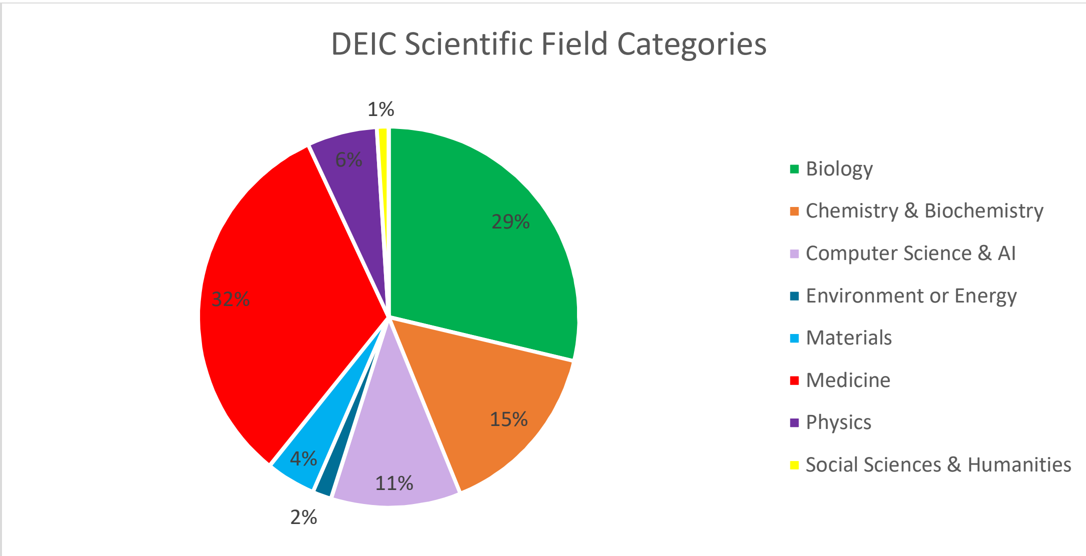


Figure 11 HPC Scientific Fields

Figure 12 display the subject area distribution for the 277 ABACUS2.0 Publications. Most publications belong to the Natural Sciences (63.6%). 86 of ABACUS2.0 publications belong to more than one subject area.

Figure 13 shows the distribution of Computerome publications, which mostly have publications from the Natural sciences and Medical sciences. For example, the single publication belonging to the humanities is also classified as being from the natural sciences.

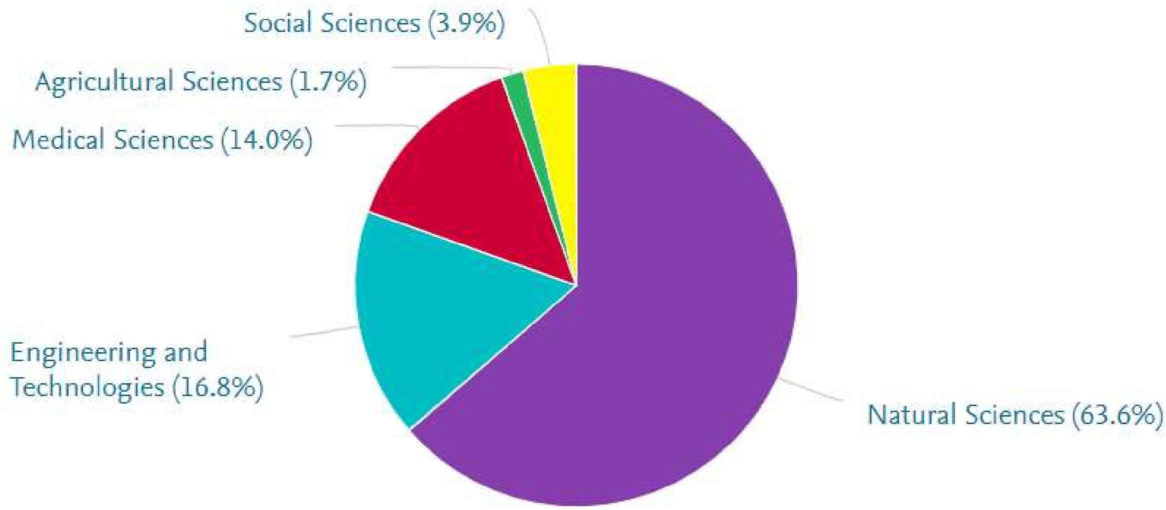


Figure 12 ABACUS2.0 FORD subject areas

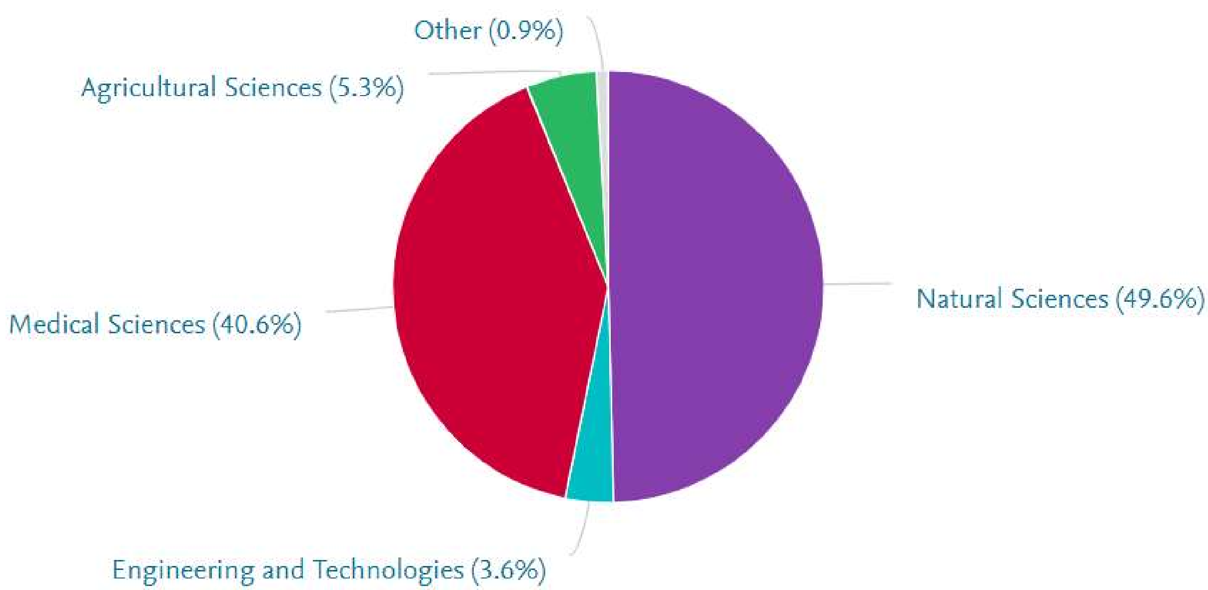


Figure 13 Computerome FORD subject areas

HPC research in comparison with research from Denmark and OECD countries

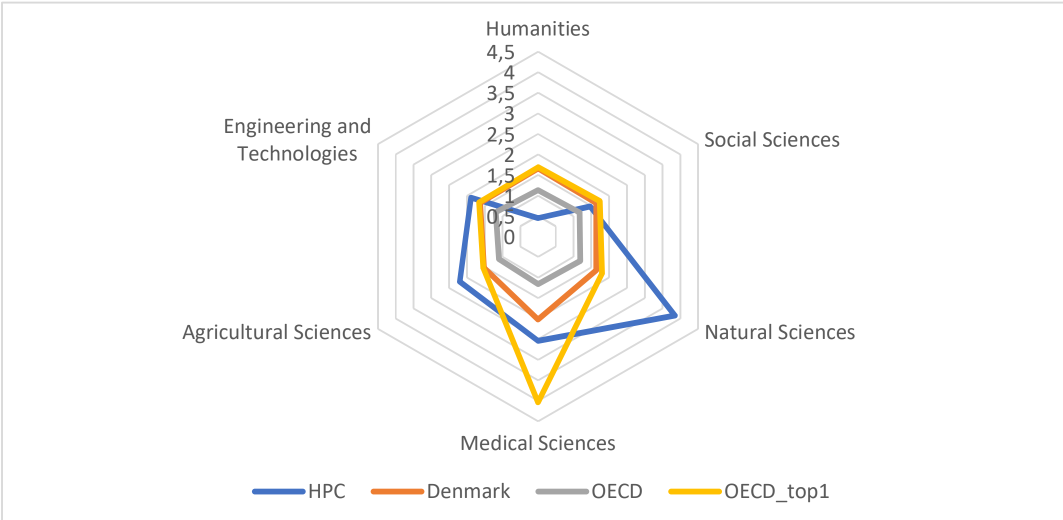


Figure 14 FWCI for HPC, Denmark, Nordic countries using the FORD – Field of Science and Technology classification (period 2014-2019 for Denmark and Nordic Countries, 2015-2019 for HPC)

The radar graph (figure 14) and table 3 show how HPC research are doing compared to research from Denmark or OECD countries. Danish research measured by FWCI are among the top5 of the OECD countries in five out of six of the subject areas during the time period 2015-2019 (Natural Sciences, Engineering and Technologies, Agricultural Sciences, Social Sciences and Humanities). In comparison, HPC publications have a higher FWCI in three out of six subject areas than OECD_1. The table and radar graph clearly show the lower usage of HPC in the humanities and Social sciences. However, for the other four subject areas where the usage of HPC are more frequent, it shows a positive impact on the publications' performance.

Table 3 FWCI for HPC, OECD, and OECD_top1 (OECD country with the highest FWCI)

	HPC		DENMARK		OECD		OECD_TOP1		
	FWCI	N	FWCI	N	FWCI	N	FWCI	N	CU
HUMANITIES	0.45	1	1.66	6,369	1.13	549,819	1.69	15,936	NL
SOCIAL SCIENCES	1.46	19	1.63	25,520	1.17	1,822,526	1.73	68,775	NL
NATURAL SCIENCES	3.85	578	1.64	84,303	1.19	5,527,216	1.8	5,407	IS
MEDICAL SCIENCES	2.54	334	2.02	65,928	1.16	3,814,090	4.04	3,684	EE
AGRICULTURAL SCIENCES	2.2	43	1.52	12,429	1.1	657,776	1.54	21,463	NL
ENGINEERING AND TECHNOLOGIES	1.89	86	1.63	32,607	1.19	2,635,764	1.66	3,055	LU
OVERALL	3.74	805	1.81	159,893	1.17	10,495,167	-	-	-

Figure 15 shows FWCI for HPC publications in relation to the FWCI for Denmark and OECD countries, as well as the “Global average”, thus the expected global average for similar publications is 1 (see also the method section). OECD countries have a FWCI 0.17 over the expected. Danish research publications do generally well in bibliometric-based evaluations, and they have a FWCI at 1.81. The HPC publications have overall a high impact and the FWCI is 3.74.

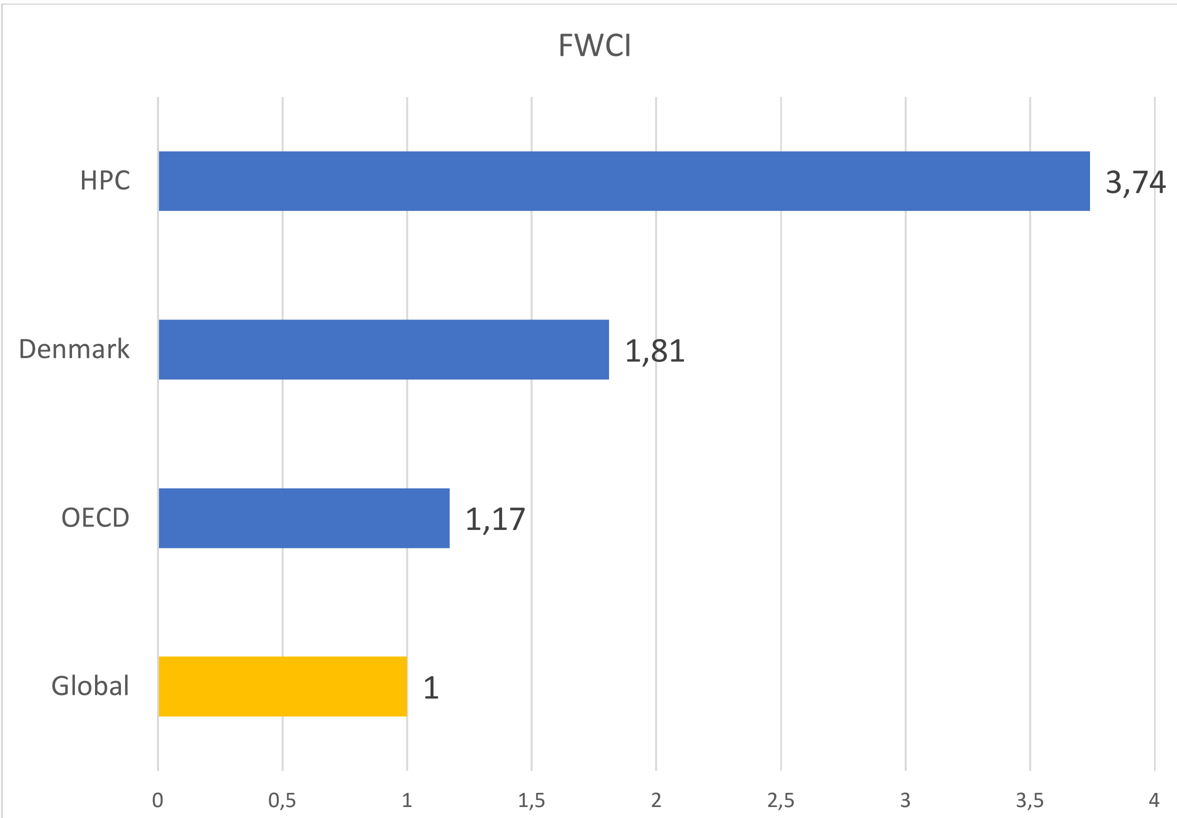


Figure 15 FWCI for publications from HPC, Denmark and OECD countries in relation to the “Global average”

Collaboration tendencies

This report examines collaboration tendencies of the HPC publications using co-authorship. Co-authorship can be measured at several levels: international, national or interinstitutional, as well as corporate or academic authorship. Table 4 shows a summary of the extent of the different types of collaboration and how much the different kinds of publications are cited.

Most publications are international co-authored (68.1%), and especially Computerome publications are created by large authorship groups (μ = 26.5 authors), fitting into the contemporary trend of medical and natural sciences working in large international research project. The mean number of citations are much higher for the international co-authored publications than national or institutional co-authored publications (see table 4).

13.4% of HPC publications are a result of a collaboration with a corporation. The corporate Computerome publications display a high tendency to be cited, while the ABACUS2.0 corporate publications have a much lower citation frequency.

Table 4 Collaboration

Type of collaboration	Collaboration	Share of publications	μ Authors	N	Citations	μ citations	FWCI
International	HPC	68.1%	20	548	20,781	37.9	4.69
	ABACUS2.0	64.0%	6.5	178	2,071	11.6	1.78
	Computerome	70.2%	26.5	370	18,710	50.6	6.1
Only national	HPC	16.5%	10.1	133	2,135	16.1	1.91
	ABACUS2.0	8.6%	5.1	24	237	9.9	1.05
	Computerome	20.7%	11.3	109	1,898	17.4	2.1
Only institutional	HPC	14.9%	4.7	120	1,578	13.1	1.53
	ABACUS2.0	25.9%	3.8	72	884	12.3	1.49
	Computerome	9.1%	6	48	694	14.5	1.58
Single authorship	HPC	0.5%	1	4	10	2.5	0.55
	ABACUS2.0	1.4%	1	4	10	2.5	0.55
Corporate	HPC	13.4%	50.3	108	4,090	37.9	5.96
	ABACUS2.0	5.0%	3	14	83	5.9	1.17
	Computerome	17.8%	47.5	94	4,007	42.6	6.67
Academic	HPC	86.6%	15.3	697	20,414	37.9	3.4
	ABACUS2.0	95.0%	5.6	264	3,119	11.8	1.65
	Computerome	82.2%	20.6	433	17,295	39.9	4.47

Country and Institution collaboration

The HPC publications have authors from 110 different countries in the byline. Figure 16 display these countries and the number of publications per country. Each country is only counted once on the publications, even if there are multiple authors affiliated with the country's institutions. Thus, if a publication has four authors, one from Denmark, two from Spain and one from Australia, each of these countries would get one credit for the publication.

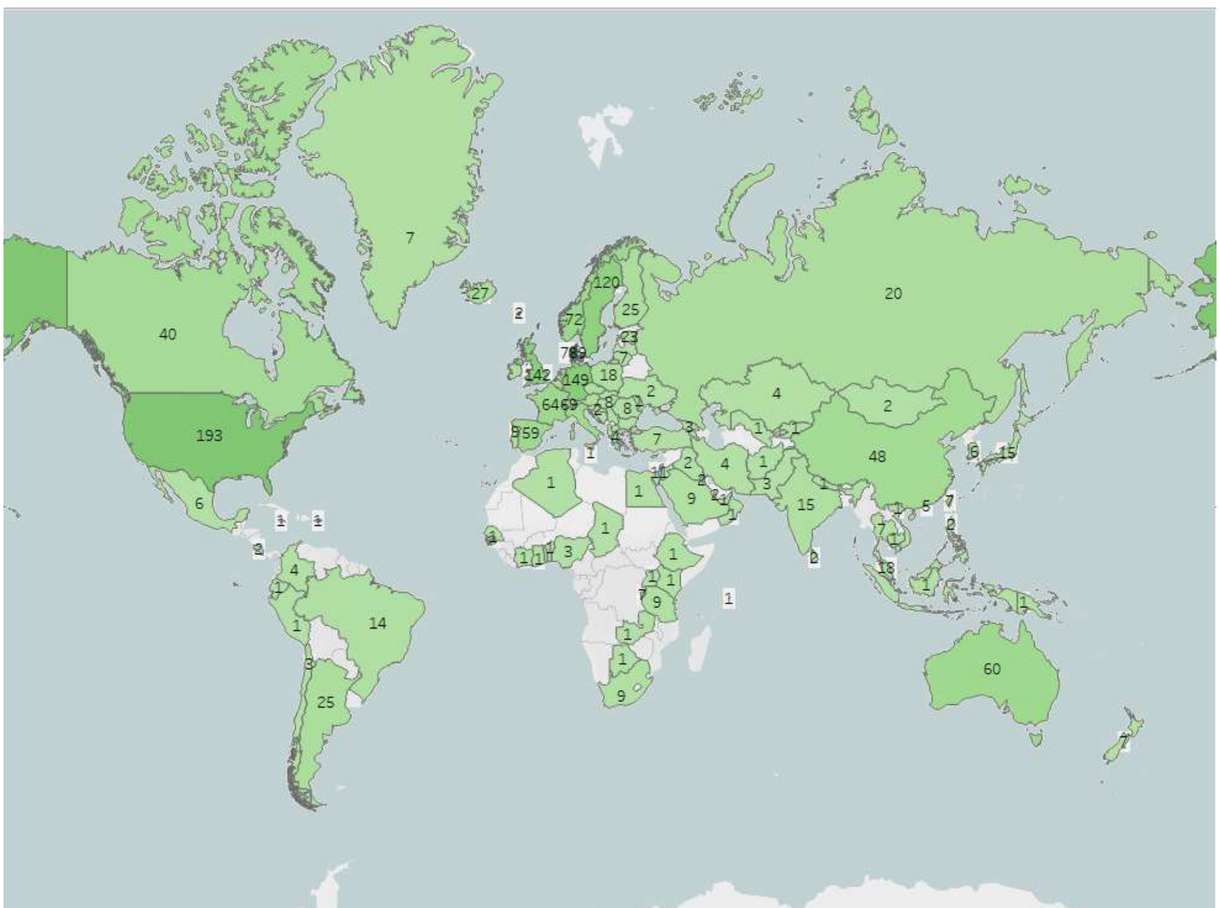


Figure 16 Country collaboration (Number of publications of the 805 DK-HPC publications captured in Scopus)

Figure 17 shows the distribution of publications per country. All countries with less than 10 publications have been merged into a group called “Other”.

Each HPC publication have in average three countries in the affiliation byline.

Computerome publications have 3.6 countries while ABACUS2.0 publications have 2.1 countries.

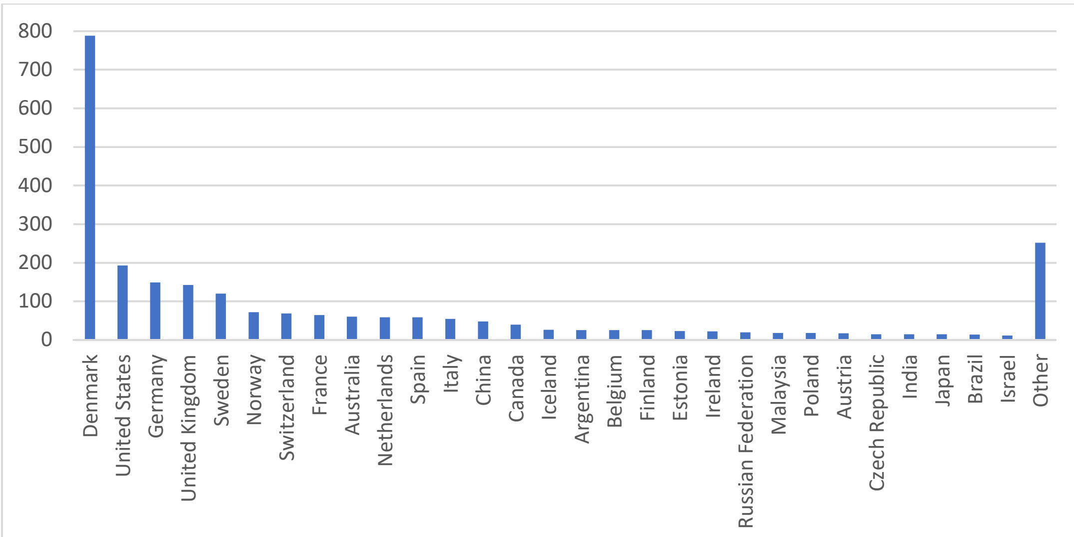


Figure 17 Number of publications per country

Topic Prominence clusters

Topic prominence clusters are dynamic in the sense that they can be large or small, grow or decline. A topic is based on a collection of documents with a common focused intellectual interest. Meaning they have been grouped together based on direct citations. So, if A cites B and C, they form topic.

Topic Clusters are formed by aggregating Topics with similar research interest together to form a broader, higher-level area of research. Scopus and SciVal have approximately 96,000 Topics, which have been matched with one of the 1,500 Topic Clusters. Unlike subject areas, a publication can only belong to one Topic and topics can only be in one Topic Cluster. Topic Clusters are formed using the same direct citation algorithm that creates the Topics.

What is prominence?

Calculating a Topic's Prominence combines 3 metrics to indicate the momentum of the Topic:

- Citation Count in year n to papers published in n and n-1
- Scopus Views Count in year n to papers published in n and n-1
- Average CiteScore for year n

Scopus view counts is the sum of abstract views in Scopus and clicks on the link in Scopus to view the full text at the publisher's website.

CiteScore is based on the average number of citations received in a calendar year by all items published in that journal in the preceding three years. Unlike the journal impact factor, Citescore includes all document types.

$$Citescore2018 = \frac{citations\ in\ 2018}{Documents\ published\ in\ 2015, 2016, 2017}$$

Overall, the prominence percentile indicates the extent of momentum of a topic cluster.

It is not a "quality" indicator.

COMP	Computer Science	MEDI	Medicine
MATH	Mathematics	PHAR	Pharmacology, Toxicology and Pharmaceutics
PHYS	Physics and Astronomy	HEAL	Health Professions
CHEM	Chemistry	NURS	Nursing
CENG	Chemical Engineering	DENT	Dentistry
MATE	Materials Science	NEUR	Neuroscience
ENGI	Engineering	ARTS	Arts and Humanities
ENER	Energy	PSYC	Psychology
ENVI	Environmental Science	SOCI	Social Sciences
EART	Earth and Planetary Sciences	BUSI	Business, Management and Accounting
AGRI	Agricultural and Biological Sciences	ECON	Economics, Econometrics and Finance
BIOC	Biochemistry, Genetics and Molecular Biology	DECI	Decision Sciences
IMMU	Immunology and Microbiology	MULT	Multidisciplinary
VETE	Veterinary		

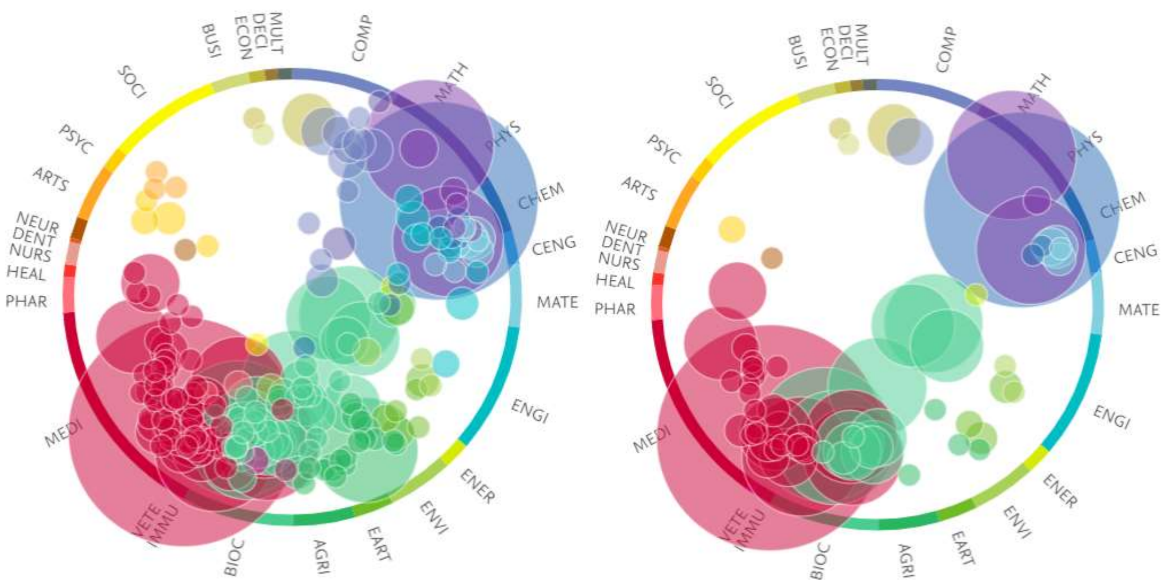


Figure 18a HPC Topic Cluster Prominence for all publications and Figure 18b for top10% prominence publications

Figure 18a displays 241 topic clusters. These topic clusters are based on 471 topics and they display a similar subject area trend as described previously in Figure 10. Thus, the 805 HPC publications are mainly a result of research from the natural sciences and medical sciences.

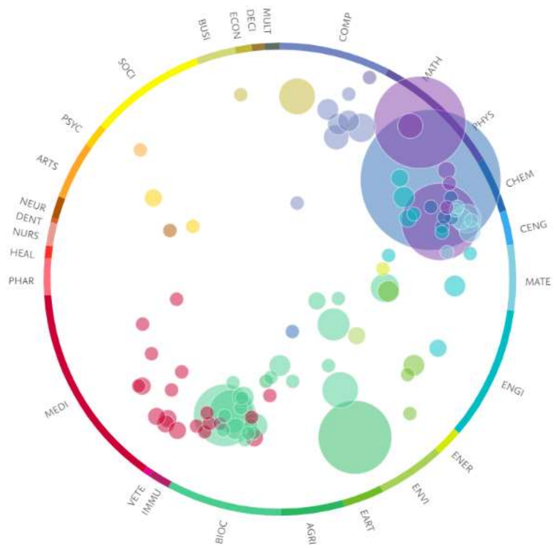
There are small clusters of research from social science and humanities (yellow and orange color).

The size of the circles is based on the number of publications, while the location of the circles is based on the topics in the cluster.

Table 5 shows the eleven largest HPC topic clusters. The two largest topic clusters in figure 18a and 18b are TC.16 (Anti-Bacterial Agents; Infection; Methicillin-Resistant Staphylococcus Aureus) with 41 publications and TC.145 (Quantum Chemistry; Density Functional Theory; Molecular Orbitals) with 35 publications. These clusters were also the largest clusters in the “Analysis of usage of ABACUS2.0 and Computerome in Danish Research” 2019 report. Figure 18b shows the 72 topic clusters belonging to the top10% topic cluster prominence (Worldwide Prominence percentile > 90). Nine of the largest topic clusters from figure 18a are in the top10% topic prominence percentile).

Table 5 Top 11 largest Topic Clusters from figure 17a

TOPIC CLUSTER (TC)	TC NUMBER	N	SHARE OF TC	FWCI	WORLDWIDE PROMINENCE PERCENTILE
ANTI-BACTERIAL AGENTS; INFECTION; METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	TC.16	41	6%	3.33	98.26
QUANTUM CHEMISTRY; DENSITY FUNCTIONAL THEORY; MOLECULAR ORBITALS	TC.145	35	16%	1.11	91.365
DNA METHYLATION; EPIGENOMICS; NEOPLASMS	TC.478	23	12%	2.59	91.968
DECAY; QUARKS; NEUTRINOS	TC.6	22	4%	3.32	98.394
T-LYMPHOCYTES; NEOPLASMS; IMMUNOTHERAPY	TC.12	20	3%	6.11	99.665
MICRORNAS; LONG UNTRANSLATED RNA; NEOPLASMS	TC.219	19	3%	1.23	99.398
PLASMONS; METAMATERIALS; SURFACE PLASMON RESONANCE	TC.47	18	2%	3.35	99.531
BIRDS; NESTS; SEABIRDS	TC.41	17	6%	1.69	89.759
PROTEINS; MOLECULAR DYNAMICS SIMULATION; MOLECULAR DYNAMICS	TC.108	16	6%	1.86	94.846
GENOME; NEOPLASMS; GENES	TC.436	16	7%	2.41	94.043
SALMONELLA; ESCHERICHIA COLI; LISTERIA MONOCYTOGENES	TC.40	16	4%	2.51	93.641



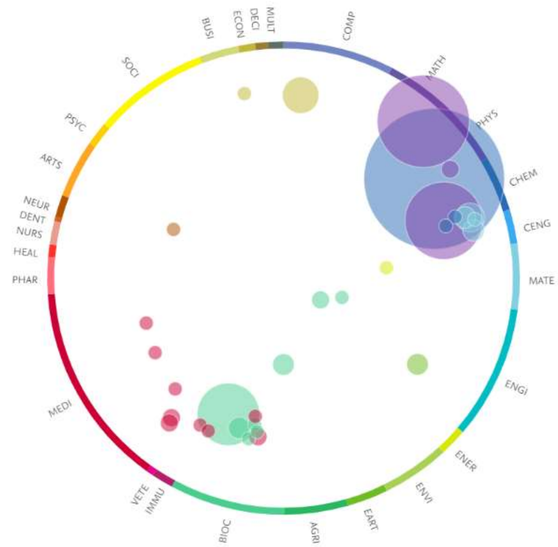


Figure 19a ABACUS2.0 prominence all and, *figure 19b ABACUS2.0 prominence top10%*

Figure 19a displays 93 topic clusters based on 139 topics, while figure 19b display 32 topics clusters. The topic clusters belong to the natural, biological, medical and social sciences. The figures show that ABACUS2.0 publications mostly are from the natural sciences (purple/blue clusters).

Figure 19a have seven topic clusters, which to some degree belongs to the social sciences; 7 publications in TC.54 (Models; Risks; Finance), 2 in TC.760 (Intelligence; Child; Psychology), 1 in TC.21 (Monetary Policy; Economic Growth; Exports), 1 in TC.928 (Testosterone; Face; Behavior), 1 in TC.10 (Attention; Brain; Learning), 1 in TC.81 Electricity; Energy; Economics, and 1 in TC.152 (Kant; Theory; Epistemic).

Table 6 displays the eleven largest topic clusters for the ABACUS2.0 publications. Five of these clusters are in the top10% prominence cluster percentile. Ten of the largest topic clusters are the same as in the "Analysis of usage of ABACUS2.0 and Computerome in Danish Research" 2019 report.

Table 6 ABACUS2.0 11 largest topic clusters

TOPIC CLUSTER (TC)	TC NUMBER	N	SHARE OF TC	FWCI	WORLDWIDE PROMINENCE PERCENTILE
QUANTUM CHEMISTRY; DENSITY FUNCTIONAL THEORY; MOLECULAR ORBITALS	TC.145	35	16%	1.11	91.365
DECAY; QUARKS; NEUTRINOS	TC.6	22	4%	3.32	98.394
PLASMONS; METAMATERIALS; SURFACE PLASMON RESONANCE	TC.47	18	2%	3.35	99.531
BIRDS; NESTS; SEABIRDS	TC.41	17	6%	1.69	89.759
DNA METHYLATION; EPIGENOMICS; NEOPLASMS	TC.478	14	7%	0.85	91.968
CALCIUM; CALCIUM SIGNALING; ION CHANNELS	TC.232	9	7%	1.05	78.112
MODELS; RISKS; FINANCE	TC.54	7	1%	0.42	90.027
PHOTOSYSTEM II PROTEIN COMPLEX; PHOTOSYNTHESIS; CHLOROPHYLL	TC.329	7	6%	1.21	83.668
NUCLEOSIDES; OLIGONUCLEOTIDES; DNA	TC.414	6	10%	0.43	63.119
ORGANIC LIGHT EMITTING DIODES (OLED); SOLAR CELLS; CONJUGATED POLYMERS	TC.61	5	1%	1.26	99.598
OPTIMIZATION; ALGORITHMS; EVOLUTIONARY ALGORITHMS	TC.259	5	2%	0.84	82.195

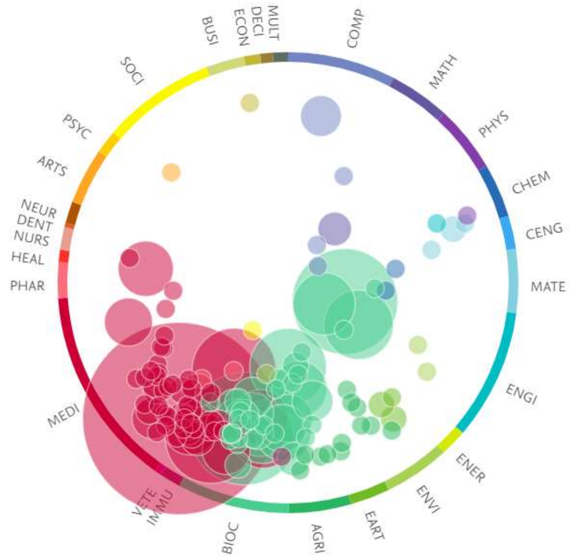


Figure 20a Computerome all and,

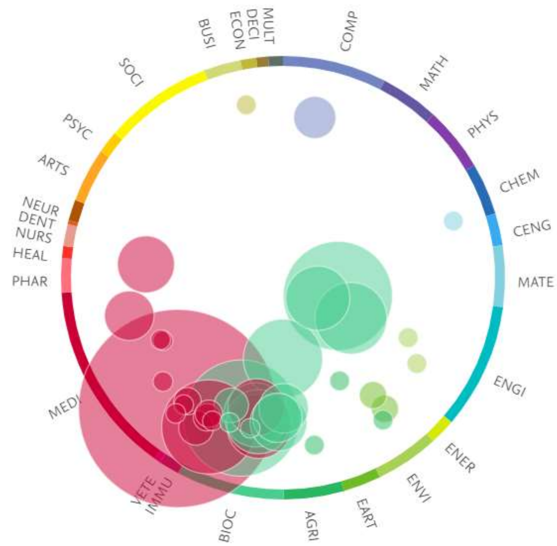


Figure 20b Computerome top10% prominence

There are 175 topic clusters in figure 20a with 39 topic clusters belonging to the top10% topic prominence percentile (Figure 20b). These topic clusters are based on 342 topics. The topic clusters are mainly from the biological and medical sciences (red and green bubbles). In comparison with figure 13 subject area, the natural sciences are not as visible. This is based on the different classifications, where biological sciences are a subcategory to natural sciences.

Figure 20a have seven topic clusters, which to some degree belongs to the social sciences; TC.677 (China; Chinese; Qing), TC.127 (Fossils; Pleistocene; Paleolithic), TC.24 (Industry; Innovation; Entrepreneurship), TC.760 (Intelligence; Child; Psychology), TC.88 (Language; Reading; Semantics), TC.338 (Israel; Text; Ancient), and TC.99 (Child; Adolescent; Schools). There is only one publication from Computerome in each of these clusters, except for TC.99, which has two publications.

Table 7 shows the 12 largest topic clusters for Computerome. Eleven out of twelve of these topic clusters belongs to the top10% topic prominence percentile. Eight of these twelve topic clusters are the same as in the "Analysis of usage of ABACUS2.0 and Computerome in Danish Research" 2019 report.

Table 7 Computerome 12 largest topic clusters

TOPIC CLUSTER (TC)	TC NUMBER	N	SHARE OF TC	FWCI	WORLDWIDE PROMINENCE PERCENTILE
ANTI-BACTERIAL AGENTS; INFECTION; METHICILLIN-RESISTANT STAPHYLOCOCCUS AUREUS	TC.16	39	6%	3.44	98.26
T-LYMPHOCYTES; NEOPLASMS; IMMUNOTHERAPY	TC.12	20	3%	6.11	99.665
MICRORNAS; LONG UNTRANSLATED RNA; NEOPLASMS	TC.219	16	2%	1.04	99.398
PROTEINS; MOLECULAR DYNAMICS SIMULATION; MOLECULAR DYNAMICS	TC.108	15	5%	1.84	94.846
DNA; MICROSATELLITE REPEATS; POPULATION	TC.604	15	15%	4.16	68.608
SALMONELLA; ESCHERICHIA COLI; LISTERIA MONOCYTOGENES	TC.40	14	4%	2.6	93.641
METAGENOME; PROBIOTICS; BACTERIA	TC.215	14	3%	5.24	98.661
GENOME; NEOPLASMS; GENES	TC.436	13	6%	2.41	94.043
ATRIAL FIBRILLATION; PATIENTS; CATHETER ABLATION	TC.25	11	2%	2.39	94.913
SCHIZOPHRENIA; PSYCHOTIC DISORDERS; ANTIPSYCHOTIC AGENTS	TC.80	11	4%	16.19	92.236
PROTEOMICS; MASS SPECTROMETRY; PROTEINS	TC.227	11	5%	6.86	92.503
GENES; GENE REGULATORY NETWORKS; GENE EXPRESSION	TC.362	11	5%	58.08	90.495

Open Access

Scopus index publications as OA if they are "Gold" or "Hybrid" OA, meaning that the publication is available for free at the journal's website. 61% of the HPC articles are registered as Open Access (OA) in Scopus.

In comparison, the Open Access Indicator for Danish research publications from 2016, 2017 and 2018 show that 36%, 45% and 54% of Danish publications are open access (https://www.oaindikator.dk/en/data?historic=yes).

However, the Danish Open Access Indicator includes publications, which are "Green" OA, meaning that publications which are "Self-archive" at the universities' repositories are include.

Thus, the Danish Open Access Indicator include a broader spectra of OA publications, than what we can detect in Scopus.

We matched the HPC publication data using DOI to the available data on the DK OA Indicator webpage (June 2020), though it did require some data cleaning, since the registration of the DOIs are not standardized. Furthermore, the 2019 data are not yet available on the website. 96.0% of the HPC publications was matched to the DK OA Indicator data.

Table 8 displays the distribution of OA publications each year. The indexed publications from 2018 display the greatest differences with the DK OA indicator registering 78.8% of the publications to be OA, while in Scopus it is 63.6% of the publications. Thus, the increasing effort and focus in Denmark to use Green OA could be the reason for the rising share of OA publications, and why it is perhaps not as visible in the Scopus database, which register Golden or hybrid OA publications.

Table 8 the distribution of OA publications each year using Scopus or Danish OA Indicator data

	SCOPUS			DK OA INDICATOR		
YEAR	Other	OA	Share of OA	Other	OA	Share of OA
2015	25	35	58.3%	32	25	43.9%
2016	42	48	53.3%	39	51	56.7%
2017	76	145	65.6%	68	143	67.8%
2018	83	145	63.6%	46	171	78.8%
2019	89	116	56.6%	-	-	-

Table 9 and figure 21 display the distribution of OA and non-OA publication types. They show how most HPC publications are articles (90%) and many publications are OA.

Table 9 distribution of OA and non-OA publication types

	Publication type	N	Percentage
OA:	Articles	452	56.1%
	Reviews	19	2.4%
	Other publications	19	2.4%
Other:	Articles	272	33.8%
	Reviews	13	1.6%
	Other publications	30	3.7%
		805	100.0%

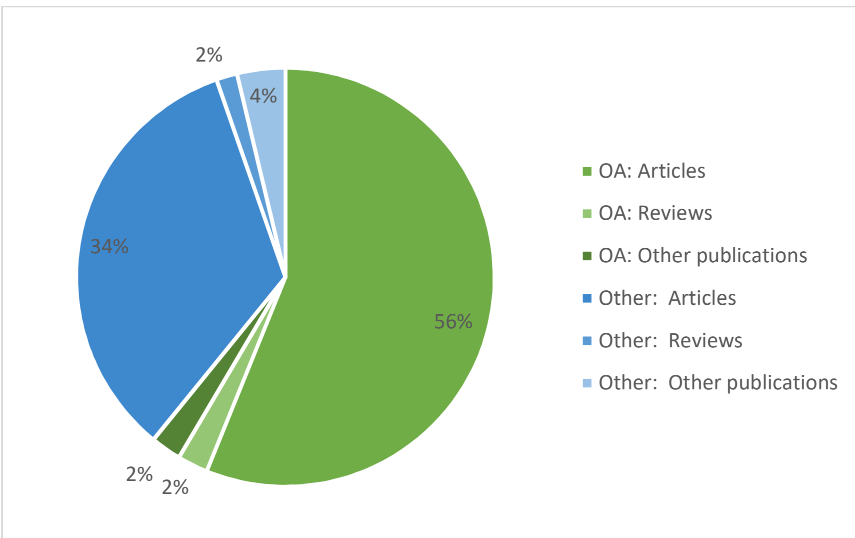


Figure 21 distribution of OA and non-OA publication types

The Societal Relevance and Societal Impact of HPC Research

We wanted to explore how the HPC publications contributes to the United Nations' (UN) Sustainable Development Goals (SDG). This is not an easy task, since there has not been developed an algorithm, which can perfectly assess and categorize the research publications as belonging to one or more of the SDGs. We also did have the resources and expertise to manually assess whether the publications could be categorized being in the subject areas of one of the SDGs. Instead we had to rely on sets of search queries developed by Elsevier (Jayabalasingham, Boverhof, Agnew, & Klein, 2019) or by bibliometricians in the Aurora network (https://aurora-network.global/project/sdg-analysis-bibliometrics-relevance/). Both sets of search queries were developed for the Scopus database, and we will refer to them as "SDG E" for Elsevier's queries and "SDG A" for the Aurora network's queries.

These queries are not perfect and are still in development. Table 10 shows how these queries gives different results when performed in the Scopus Database. The SDG search queries was combined with queries containing the publications' DOI or PMID for either ABACUS2.0 or Computerome. In total we matched 349 publications to an SDG goal. Five publications were matched to two SDG goals. Table 11 shows how these SDGs matches the assigned DEIC Scientific Fields Categories. The publications matching SDG3 belongs mostly to DEIC's Scientific Fields' "Biology" and "Medicine", which could indicate that these publications match the category "Good Health and Well-being".

Still, there were very few results, which could indicate that the research performed using ABACUS2.0 or Computerome do not aligned with the UN's SDGs. However, this could also be an indication that the search queries need more optimization before representing research publications matching the SDGs. Armitage, Lorenz, and Mikki (2020) have criticizes SDG E queries and the usage of them in rankings and evaluations, since it is too "*premature to trust commercial SDG-analyses for anything other than exploratory purposes at this stage in their development*". Therefore, the results in table should be interpret with cautions.

Table 10

SDG	SDG Name	SDG E	SDG E AND ABACUS2.0	SDG E AND Computerome	SDG E AND HPC	SDG A	SDG A AND ABACUS2.0	SDG A AND Computerome	SDG A AND HPC
1	No Poverty	12,927	0	0	0	43,005	0	0	0
2	Zero Hunger	107,562	0	1	1	22,291	0	0	0
3	Good Health and Well-being	3,716,799	46	293	339	81,063	1	2	3
4	Quality Education	29,305	0	0	0	52,163	0	1	1
5	Gender Equality	40,700	0	0	0	22,068	0	0	0
6	Clean Water and Sanitation	52,271	0	0	0	96,257	0	5	5
7	Affordable and Clean Energy	436,184	2	0	2	47,700	0	0	0
8	Decent Work and Economic Growth	103,825	0	0	0	32,314	0	0	0
9	Industry, Innovation and Infrastructure	45,096	0	0	0	26,137	0	0	0
10	Reduced Inequality	53,730	0	0	0	18,039	0	0	0
11	Sustainable Cities and Communities	162,550	0	2	2	96,570	0	1	1
12	Responsible Consumption and Production	97,496	0	0	0	176,231	1	2	3
13	Climate Action	204,999	2	4	6	327,436	3	7	10
14	Life Below Water	119,129	0	0	0	145,717	0	3	3
15	Life on Land	126,269	2	2	4	Not working	-	-	-
16	Peace and Justice Strong Institutions	191,662	0	0	0	127,445	0	1	1
17	Partnerships to achieve the Goal	Not created	-	-	-	22,602	0	0	0

Table 11

DEIC SCIENTIFIC FIELDS	SDG2	SDG3	SDG7	SDG11	SDG13	SDG15	NO SDG MATCH	TOTAL
BIOLOGY		92		1	1		137	231
CHEMISTRY & BIOCHEMISTRY		21					101	122
COMPUTER SCIENCE & AI		20	1		1		68	89
ENVIRONMENT OR ENERGY		1	1	2	3	1	6	13
MATERIALS		1					33	34
MEDICINE	1	201		1	1		58	260
PHYSICS		3					45	48
SOCIAL SCIENCES & HUMANITIES							8	8
TOTAL	1	339	2	4	6	1	456	805

Final remarks

The report shows that HPC publications are mostly published in higher impact journals according to the three journal rankings. The HPC publications have in average received more citation than average Danish research publications in the time period 2015-2020 (June). Still, we need a longer time window for the publications published in 2017-2019. Thus, we will need to follow the publishing trend of the HPC research projects over a longer time period to better show the impact of the research.

References

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. Advance Publication.* doi:10.1162/qss_a_00071

The Danish Open Access Indicator. (2020). Retrieved from https://www.oaindikator.dk/en

Jayabalasingham, B., Boverhof, R., Agnew, K., & Klein, L. (2019). *Identifying research supporting the United Nations Sustainable Development Goals*.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics, 7*(2), 272-285. doi:https://doi.org/10.1016/j.joi.2012.11.011