Information Research - Vol. 30 No. 1 (2025)

Comparing the SDG-classification systems of Dimensions, InCites and SciVal for the University of Helsinki

Petri Turunen and Terhi Sandgren
DOI: https://doi.org/10.47989/ir30140516

Abstract

Introduction. We determined how the publications of the University of Helsinki were attributed to the United Nations sustainable development goals (SDGs) in three different databases. The databases were Dimensions, Scopus and Web of Science. We also looked at the overlaps.

Method. DOIs for the publications were collected from our current research information system. These DOIs were then used to create publication sets in the databases. SDG-attributions of publications in these publication sets were collected for analysis.

Analysis. Total publication counts and overlaps were calculated with simple Excel functions. This was done for all publications and separately for only those publications that were found in all three databases. Venn graphs were plotted for each sustainable development goal.

Results. The analysis revealed significant differences between the attributions. Overlaps tended to be low usually covering only a few percent up to, at best, half of the publications. Total number of publications also varied with manyfold differences. The attributions were also found to be rather unstable due to updates in attribution systems.

Conclusions. Sustainable development goal attributions are very database dependent and remain unstable. At present, there is no responsible way to track them for evaluative purposes. It is unlikely that the situation will change in the future. This is due to their general scope and a lack of stabilising mechanisms.

Introduction

The United Nations 2030 agenda for sustainable development launched in 2015 and introduced 17 sustainable development goals (SDGs) (Sachs 2012; UN 2015). Ever since then, Universities have been interested in evaluating how their research relates to these goals (for examples see Al-Raeei, 2023; Nature, 2021; Pakkan, et al., 2023; Repiso, et al., 2023). In University of Helsinki we are developing ways of monitoring and analysing the share of publications related to sustainable development.

Several analytical tools for tracking SDG contributions have cropped up in recent years (see e.g., Borchardt, et al., 2022; Confraria, et al., 2024; Hajikhani and Suominen, 2022; Our World in Data team, 2023; Rodenburg, et al., 2021). Major citation databases such as Scopus, Web of Science and Dimensions now come with SDG-information. However, these databases tend to use different methods for relating publications under various SDGs.

Scopus based analytical database SciVal uses AI-enhanced keyword searches. There have been many iterations on this approach. The first published search queries were presented in 2019 and contained relatively simple search strings (Jayabalasingham, et al., 2019). The 2021 queries saw the introduction of machine learning techniques (Rivest, et al., 2021). 2022 and 2023 queries introduced further refinements (Bedard-Vallee, et al., 2023; Roberge, et al., 2022).

Web of Science based analytical database InCites uses citation-based clusters (Garcia, 2022). There is a set of initial keyword searches, which are then used to identify micro citation topics. If a citation topic is strongly associated with an SDG, then all publications in that citation topic will be attributed to that SDG. This method was originally introduced by the STRINGS project (see Confraria et al., 2024 for details).

The Dimensions database uses curated machine learning approaches (Wastl, et al., 2020). Curated data was fed to a machine learning algorithm to generate an automatic classification model. This approach also incorporated natural language processing and a quality control process.

The problem is that different approaches tend to yield different results (Armitage, et al., 2020; Purnell, 2022; Raman, et al., 2023). All classification contains an interpretative element, but since SDGs are related to important policy goals and come with the promise of numerical indicators, this case of ambiguity is more troublesome than most. SDGs are also very general in their scope and do not benefit from unifying mechanisms such as having a long historical tradition like traditional disciplines. The connection between SDG-research and SDG-related practical solutions is also strained at best (Ludwig, et al., 2022). This lack of an external cohesion makes it very difficult to evaluate the different SDG-mapping methods. In fact, it has been suggested that one should not assume there is some correct mapping method and instead just interpret the different methods as merely different perspectives (Rafols, et al., 2021).

We wanted to find out *what the differences in SDG-mapping methods look like at the level of a local institution.* Our target institution, our university, has a very broad range of research fields and a well-maintained current research information system. The University of Helsinki has roughly 31,000 degree students, 4,500 teaching and research staff, around 9,000 peer-reviewed publications per year and faculties in agriculture and forestry, biological sciences, humanities, education sciences, medicine, natural sciences and mathematics, law, pharmacy, social sciences, theology and veterinary sciences. Our institution allows for a broad evaluation of SDG-attributions over a range of research topics with accurate coverage information. The more local focus also allows us to cover all SDG and see firsthand the kind of anomalies that can show up on a more restricted scale.

Methods

To find out our SDG distributions, we used our Pure based Current Research Information System (CRIS), Tuhat. Scientific publications in our CRIS are manually verified and researchers are required to submit their publications to Tuhat. For this reason, publication information in our CRIS is highly reliable and has high coverage.

We gathered all peer-reviewed scientific publications from the University of Helsinki between the years 2020-2022. This included articles, review articles, conference articles, peer-reviewed notes and letters as well as books and edited volumes. Our university also published 1,452 dissertations during 2020-2022. Only 15 of them had a Digital Object Identifier (DOI). Dimension contained 13 of them, while InCites and SciVal contained none. Of those 13 dissertations, only one had an SDG-attribution. Dissertations were therefore not included.

In total, there were 26,639 publications. The publications were collected in April 2023.

22,563 publications (85%) had a DOI. The publications were found in the Dimensions, InCites and SciVal databases using their respective DOIs. In InCites and Scival, the databases DOI importing feature was used. For Dimensions, we performed direct DOI-searches. InCites found 18,761 (83%) of these publications, SciVal 20,169 (89%) and Dimensions 21,365 (95%) respectively. We exported these DOI-based datasets from the databases to get the SDG-attribution information. That information was then compared on an article-to-article level. Overlaps were calculated through simple Excel functions and the formulas manually checked to eliminate possible errors.

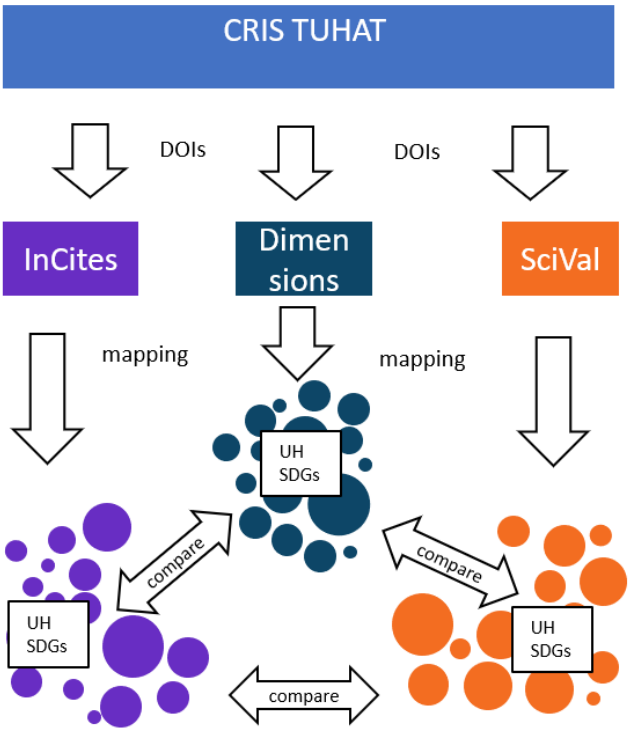

Figure 1. Extracting and comparing SDG-information from the databases.

SciVal and InCites also come with ready-made SDG-information based on affiliate information, which can be used to quickly gather SDG-information on a particular research organisation. When comparing different research organisations, this is likely the most straightforward way to proceed. However, we noticed that using such affiliation data can lead to varying results. In our testing,

SciVal found between 25%-72% of our DOI-based publication SDG-counts while InCites found between 96%-125%. So, for SciVal, a significant number of SDG-publications was left out if pure affiliation information was used instead of validated DOIs. For InCites, affiliation information gave a sizeable number of additional publications, which may not actually belong to our research organisation.

Use of additional identifiers such as WOS ID can be used to improve the coverage of the publications sets obtained from the databases.

However, this requires careful manual checking of the results and tends to result in very few additional publications for the case of our university. This is because we have a very well-maintained CRIS and hence the missing publications tend to either have wrong metadata or lack proper identifiers (DOI).

Furthermore, identifiers from all the three databases have not been comprehensibly collected into our CRIS. Use of database related identifiers could then introduce additional bias. We believe that using pure DOIs makes the analysis fairer and more transparent.

Results

For each SDG, we determined the total number of publications in each database and overlaps between each of the three SDG-attributions. There are two factors that affect the overlaps: 1) different SDG-attributions and 2) differences in database coverage. To demonstrate both factors, we created two sets of SDG values. The first set contains all SDG-publications from each database. The second set contains only those SDG-publications that are covered by all three databases (15,786 publications, marked as SDG*).

The total numbers of SDG-publications were the following:

SDG	InCites	SciVal	Dimensions	InCites*	SciVal*	Dimensions*
1	200	137	13	154	90	8
2	659	354	167	611	295	137
3	9625	4512	3167	8495	3797	2617
4	484	449	650	419	317	362
5	669	181	60	593	122	30
6	237	331	24	221	247	17
7	159	248	259	138	209	200
8	141	331	76	112	247	36
9	152	218	15	97	162	9
10	278	359	94	192	260	63
11	966	469	120	732	365	89
12	162	332	68	141	283	47
13	2359	684	673	2110	591	573
14	347	308	328	319	282	296
15	1794	701	696	1656	619	605
16	79	355	305	57	243	156

Figure 2. Number of SDG publications in InCites, SciVal and Dimensions. InCites*, SciVal* and Dimensions* only counts publications that are covered by all three databases. Databars indicate the percentage of jointly covered publications to all dataset SDG-publications in the database.

There are rather large variances between different SDG-counts. InCites has by far the highest number of University of Helsinki publications with associated SDGs, while Dimensions tends to have far fewer. For individual SDGs, InCites tends to have the highest number of articles, but there are also SDGs in which InCites has the least number of attributed publications. Dimensions has the most SDG-attributed publications for SDGs 4 and 7, but this seems to be caused by Dimensions having better overall coverage.

There are a total of seventeen SDGs, but we only considered the first sixteen. SciVal had 388 publications attributed to SDG 17, but Dimensions only had 4 and InCites had zero. Based on these numbers, SDG 17 was omitted from the analysis.

Restricting the analysis to SDG-publication that are covered by all three databases, we find that InCites and SciVal SDG-publications can be moderately affected, but Dimensions can have rather drastic reductions in their SDG-totals (over 50%). This can be due to Dimensions having better overall coverage, while also having the least total number of SDG-attributed publications.

SDG-overlaps in terms of total numbers were the following:

SDG	Only InCites	Only SciVal	Only Dimensions	InCites & SciVal	SciVal & Dimensions	InCites & Dimensions	InCites & SciVal & Dimensions
1	180	112	6	20	7	2	2
2	508	186	50	125	91	74	48
3	4785	518	688	3837	1476	2322	1319
4	278	208	384	125	185	150	69
5	633	127	27	33	30	12	9
6	184	274	11	51	11	7	5
7	102	141	146	38	94	44	25
8	115	282	41	23	32	9	6
9	139	203	7	12	7	5	4
10	218	268	48	56	42	11	7
11	776	247	37	186	79	47	43
12	126	283	32	30	30	17	11
13	1709	135	123	455	355	456	261
14	181	138	194	134	102	98	66
15	1163	151	165	474	374	455	298
16	45	260	212	27	86	25	18

Table 1. SDG-overlaps. “Only X” indicates that the publications are SDG-attributed to the particular SDG only in database X. “Database X & Database Y” indicates that the publications are SDG-attributed to the particular SDG in both Database X and Database Y. “InCites & SciVal & Dimensions” column contains publications that are attributed to the particular SDG in all three databases. Publications can have multiple SDG attributions.

SDG*	Only InCites	Only SciVal	Only Dimensions	InCites & SciVal	SciVal & Dimensions	InCites & Dimensions	InCites & SciVal & Dimensions
1	141	72	1	13	7	2	2
2	470	144	27	115	84	74	48
3	4063	269	196	3429	1418	2322	1319
4	226	124	131	112	150	150	69
5	562	77	1	28	26	12	9
6	170	192	4	49	11	7	5
7	83	114	97	36	84	44	25
8	94	216	11	15	22	9	6
9	85	148	1	11	7	5	4
10	147	186	19	41	40	11	7
11	561	170	14	167	71	47	43
12	108	240	14	27	27	17	11
13	1493	90	38	422	340	456	261
14	158	119	164	129	100	98	66
15	1045	97	82	454	366	455	298
16	29	172	81	21	68	25	18

Table 2. SDG-overlaps for publications that are covered by all three databases. "Only X" indicates that the publications are SDG-attributed to the particular SDG only in database X. "Database X & Database Y" indicates that the publications are SDG-attributed to the particular SDG in both Database X and Database Y. "InCites & SciVal & Dimensions" column contains publications that are attributed to the particular SDG in all three databases. Publications can have multiple SDG-attributions.

Of note is that, due to low overall SDG-publication counts, Dimensions occasionally has very few SDG-publications that are not similarly SDG-attributed in at least one of the two other databases. This is particularly true for SDGs 1, 5, 6 and 9 and especially if we restrict to publications that are covered by all three databases. The exceptions seem to be SDG 4, 7, 14 and 16.

The proportional overlaps were:

SDG	Shared InCites & SciVal of InCites	Shared InCites & SciVal of Scival	Shared Dimensions & InCites of InCites	Shared Dimensions & InCites of Dimensions	Shared Dimensions & SciVal of SciVal	Shared Dimensions & Scival of Dimensions	All shared of InCites	All shared of Scival	All shared of Dimensions
1	10 %	15 %	1 %	15 %	5 %	54 %	1 %	1 %	15 %
2	19 %	35 %	11 %	44 %	26 %	54 %	7 %	14 %	29 %
3	40 %	85 %	24 %	73 %	33 %	47 %	14 %	29 %	42 %
4	26 %	28 %	31 %	23 %	41 %	28 %	14 %	15 %	11 %
5	5 %	18 %	2 %	20 %	17 %	50 %	1 %	5 %	15 %
6	22 %	15 %	3 %	29 %	3 %	46 %	2 %	2 %	21 %
7	24 %	15 %	28 %	17 %	38 %	36 %	16 %	10 %	10 %
8	16 %	7 %	6 %	12 %	10 %	42 %	4 %	2 %	8 %
9	8 %	6 %	3 %	33 %	3 %	47 %	3 %	2 %	27 %
10	20 %	16 %	4 %	12 %	12 %	45 %	3 %	2 %	7 %
11	19 %	40 %	5 %	39 %	17 %	66 %	4 %	9 %	36 %
12	19 %	9 %	10 %	25 %	9 %	44 %	7 %	3 %	16 %
13	19 %	67 %	19 %	68 %	52 %	53 %	11 %	38 %	39 %
14	39 %	44 %	28 %	30 %	33 %	31 %	19 %	21 %	20 %
15	26 %	68 %	25 %	65 %	53 %	54 %	17 %	43 %	43 %
16	34 %	8 %	32 %	8 %	24 %	28 %	23 %	5 %	6 %

Figure 3. SDG-overlaps in terms of relative percentiles. “Shared database X & Database Y of Database X/Y” indicates the number of shared SDG-publications in Database X and Database Y divided by the total number of SDG-publication in Database X/Y. “All shared of Database Z” indicates the number of shared InCites, SciVal and Dimensions SDG-publications divided by the total number of SDG-publication in Database Z.

SDG	Shared InCites & SciVal of InCites	Shared InCites & SciVal of Scival	Shared Dimensions & InCites of InCites	Shared Dimensions & InCites of Dimensions	Shared Dimensions & SciVal of SciVal	Shared Dimensions & Scival of Dimensions	All shared of InCites	All shared of Scival	All shared of Dimensions
1	8 %	14 %	1 %	25 %	8 %	88 %	1 %	2 %	25 %
2	19 %	39 %	12 %	54 %	28 %	61 %	8 %	16 %	35 %
3	40 %	90 %	27 %	89 %	37 %	54 %	16 %	35 %	50 %
4	27 %	35 %	36 %	41 %	47 %	41 %	16 %	22 %	19 %
5	5 %	23 %	2 %	40 %	21 %	87 %	2 %	7 %	30 %
6	22 %	20 %	3 %	41 %	4 %	65 %	2 %	2 %	29 %
7	26 %	17 %	32 %	22 %	40 %	42 %	18 %	12 %	13 %
8	13 %	6 %	8 %	25 %	9 %	61 %	5 %	2 %	17 %
9	11 %	7 %	5 %	56 %	4 %	78 %	4 %	2 %	44 %
10	21 %	16 %	6 %	17 %	15 %	63 %	4 %	3 %	11 %
11	23 %	46 %	6 %	53 %	19 %	80 %	6 %	12 %	48 %
12	19 %	10 %	12 %	36 %	10 %	57 %	8 %	4 %	23 %
13	20 %	71 %	22 %	80 %	58 %	59 %	12 %	44 %	46 %
14	40 %	46 %	31 %	33 %	35 %	34 %	21 %	23 %	22 %
15	27 %	73 %	27 %	75 %	59 %	60 %	18 %	48 %	49 %
16	37 %	9 %	44 %	16 %	28 %	44 %	32 %	7 %	12 %

Figure 4. SDG-overlaps in terms of relative percentiles for publications that are covered by all three databases. “Shared Database X & Database Y of Database X/Y” indicates the number of shared SDG-publications in Database X and Database Y divided by the total number of SDG-publication in Database X/Y. “All shared of Database Z” indicates the number of shared InCites, SciVal and Dimensions SDG-publications divided by the total number of SDG-publication in Database Z.

An immediate observation is that InCites and SciVal have rather different SDG-attributions. At best, Scival can match only 40% of InCites SDG-publications and at worst it only matches 5%. SciVal is a bit better matched by InCites but InCites also has a lot more SDG-attributed publications in total. Dimensions is matched reasonably well by both InCites and SciVal but Dimensions itself does not cover either InCites or SciVal that well. This is probably doe to Dimensions just having

less overall SDG-publications. SDG-publications matched by all databases are low across the board, signalling that SDG-attributions are rather inconsistent.

We can also look at individual SDGs to get a better understanding of the various overlaps. We constructed Venn diagrams for the overlaps and coloured the bubbles based on database thematic colours (Dimensions = blue, InCites = violet and SciVal = orange). The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases.

SDG 1: No Poverty.

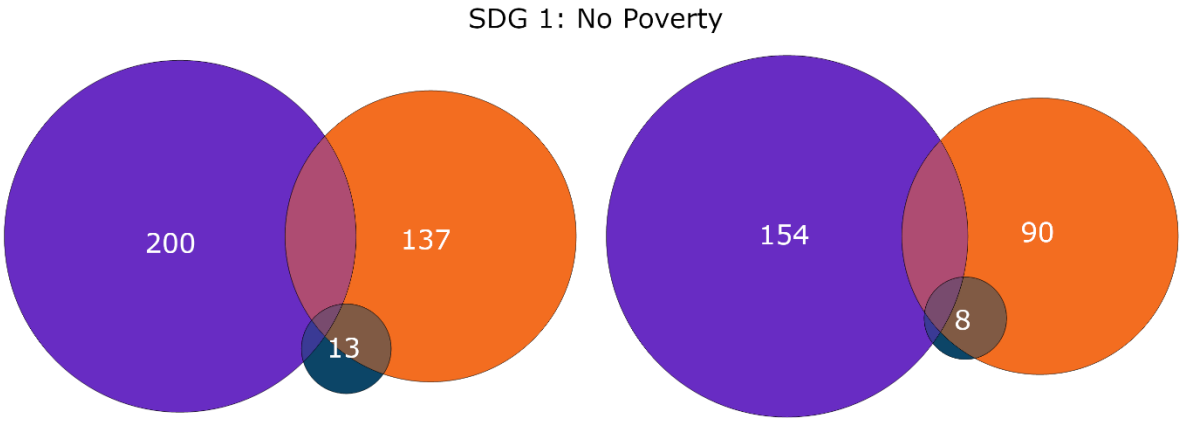

Figure 5. SDG 1: No Poverty overlaps for the three used databases. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

The first SDG 1 has low overlap between InCites and SciVal (10% and 15%). It is even a bit lower when we account for coverage differences (8% and 14%). Dimensions has far fewer publications attributed to this SDG than the other two databases and they mostly overlap with SDG-publications in SciVal (54%/88%). Nevertheless, even here the overlap is not complete.

SDG 2: Zero Hunger.

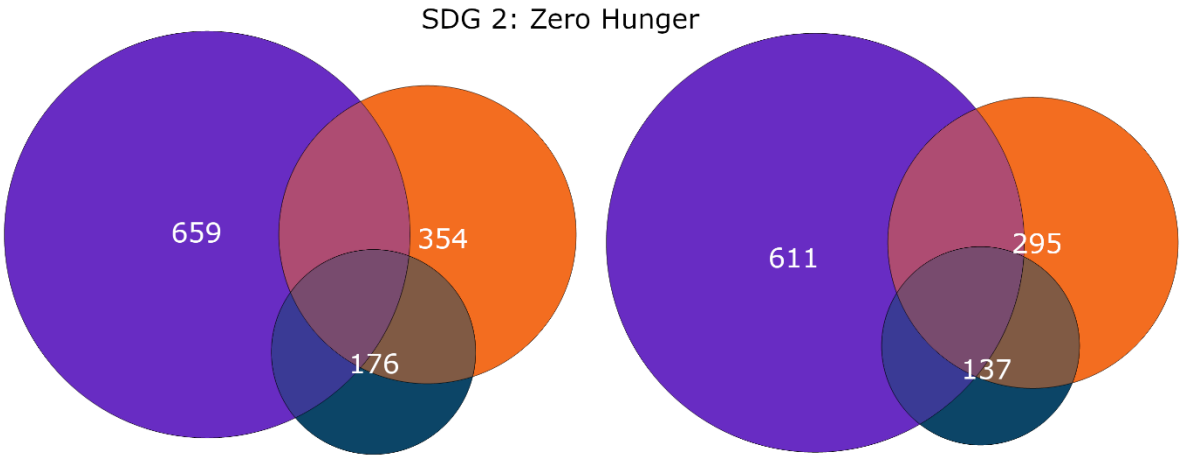

Figure 6. SDG 2: Zero Hunger. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 2 has a quite clear difference between InCites and Scival (19% and 35%/39%), but SciVal and dimensions have average overlaps (26%/28% and 54%/61%). InCites covers Dimensions rather well (44%/54%) but InCites also has a lot more SDG 2-attributed publications (659/611 compared to 167/137).

SDG 3: Good Health and Well-being.

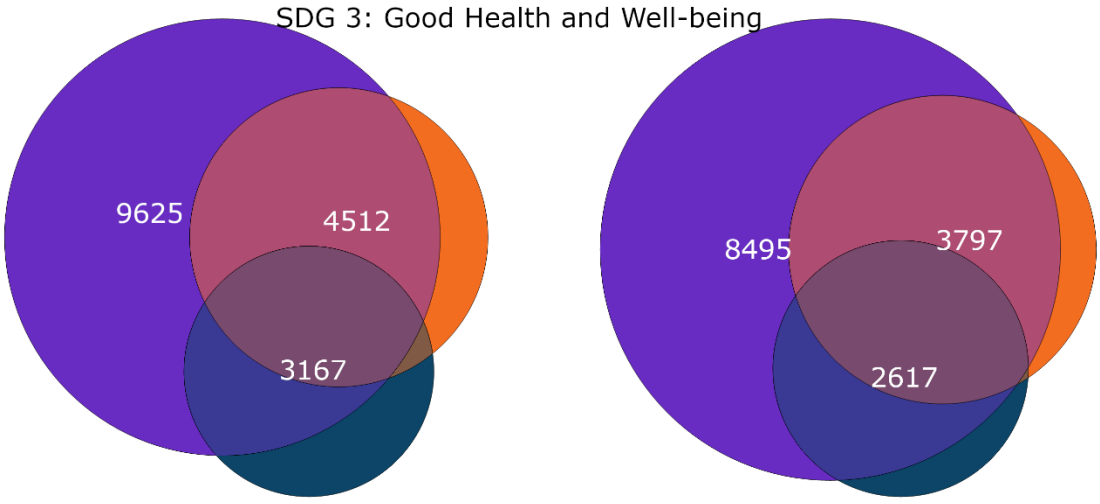

Figure 7. SDG 3: Good Health and Well-being. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 3 has by far the most overall publications (InCites: 9425/8495, SciVal: 4512/3794 and Dimensions: 3167/2717). This is because SDG 3 is of a very general nature, covering human health and welfare and therefore very closely connected to the field of medicine. Our medical faculty has high publication counts averaging around 3000 peer-reviewed publications yearly with high coverage in international publication databases. High SDG 3 publication counts were thus to be expected. InCites has a very broad take on SDG 3 and it ends up covering both SciVal and Dimensions here. However, SciVal and Dimensions only cover at most 40% of InCites publications. Due to the high publication counts, database differences here can lead to very different overall SDG counts.

SDG 4: Quality Education.

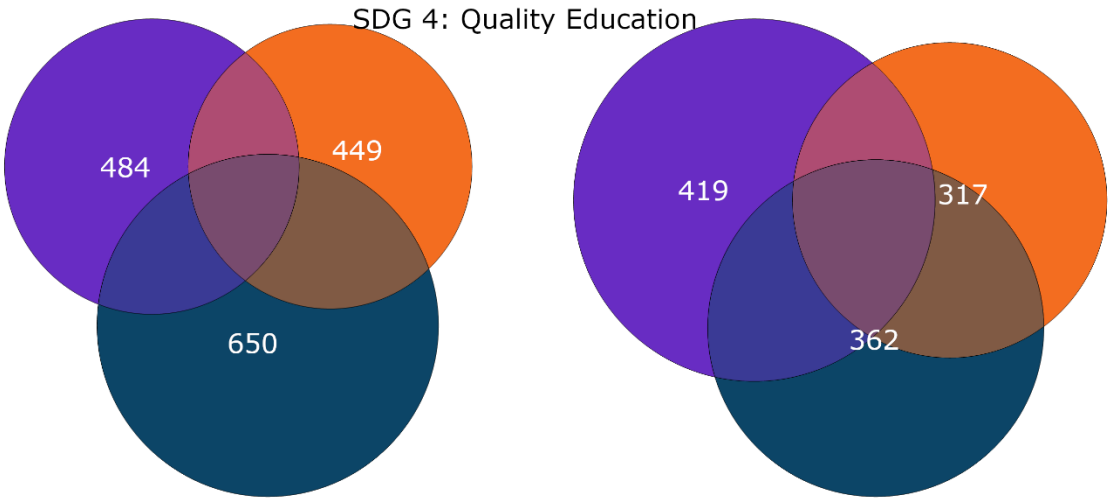

Figure 8. SDG 4: Quality Education. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 4 is evenly spread out. Each database has their own take on quality education. Overlaps are low to average (41%/47% to 23%/27%). It is worth noting that education related publications are not that well covered in InCites or Scival. Dimensions has better coverage and consequently loses proportionally more publications when we consider only publications found in all three databases. Better coverage of education research could yield more consistent results.

SDG 5: Gender Equality.

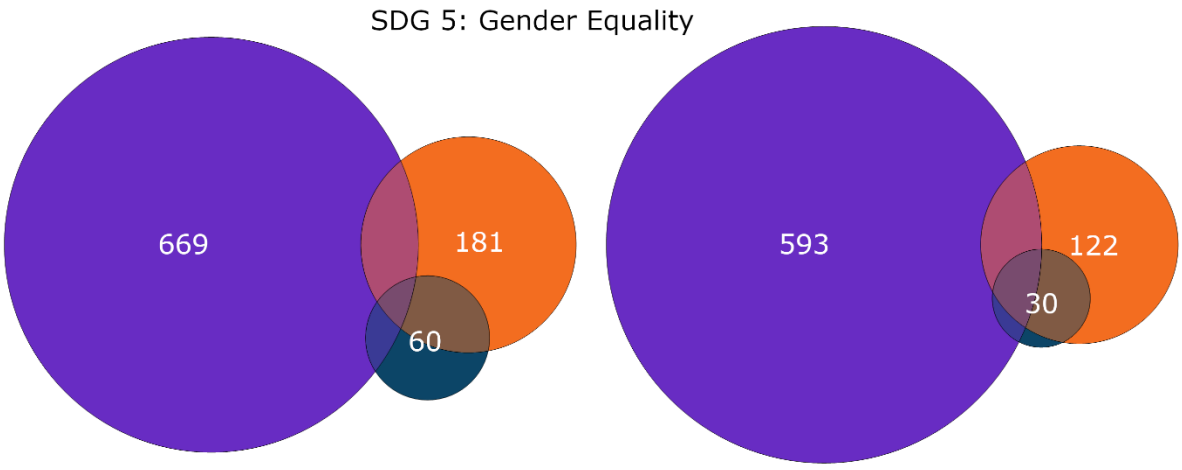

Figure 9. SDG 5: Gender Equality. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 5 has a particularly small overlap between InCites and Scival (5% and 18%/23%). SciVal and Dimensiosn have higher overlaps (50%/87% and 17%/21%). InCites also has a lot more SDG 5-attributed publications than SciVal or Dimensions. A cursory study of the publications indicates that the underlying reason seems to be that InCites attributes a lot of medical research to SDG 5 where as SciVal and Dimensions are more focused on social sciences.

SDG 6: Clean Water and Sanitation.

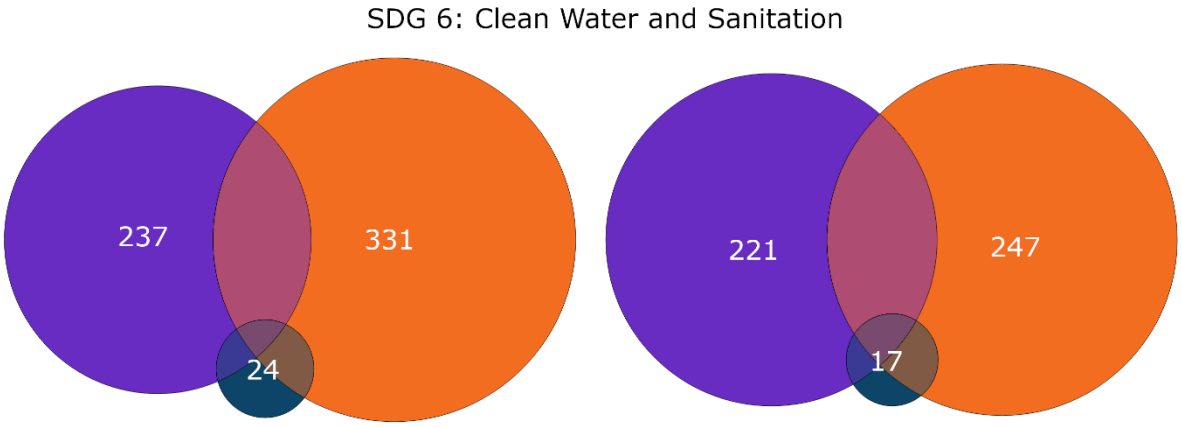

Figure 10. SDG 6: Clean Water and Sanitation. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

InCites and SciVal have rather similar publications numbers, but their overlap is low (22% and 15%/20%). Dimensions has few publications in comparison and publications overlap with both InCites and SciVal evenly.

SDG 7: Affordable and Clean Energy.

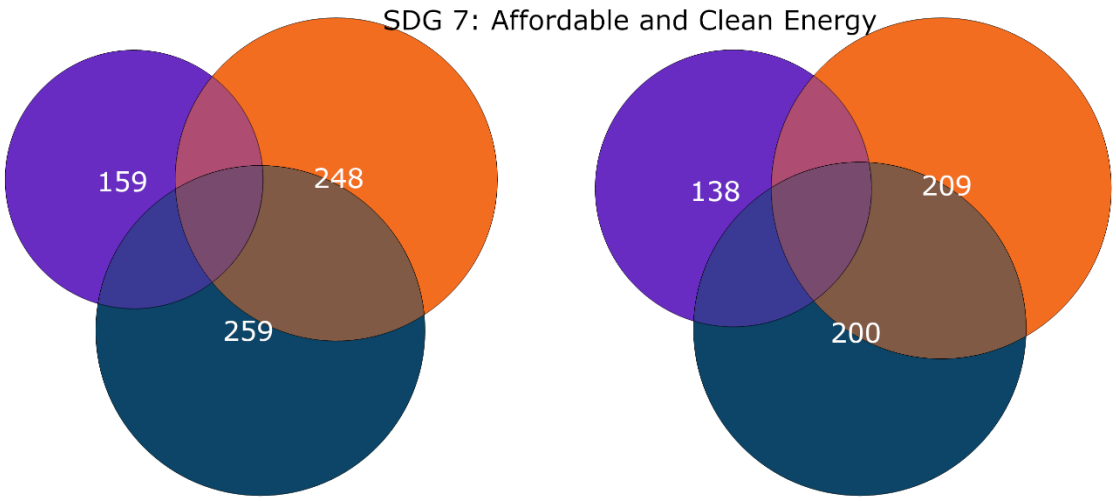

Figure 11. SDG 7: Affordable and Clean Energy. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

For SDG 7, the publication counts are similar. The overlap between Dimensions and Scival (38%/40% and 36%/42%) is a bit higher than overlaps with InCites (24%/26% and 15%/17% for SciVal and 28%/32% and 17%/22% for Dimensions).

SDG 8: Decent Work and Economic Growth.

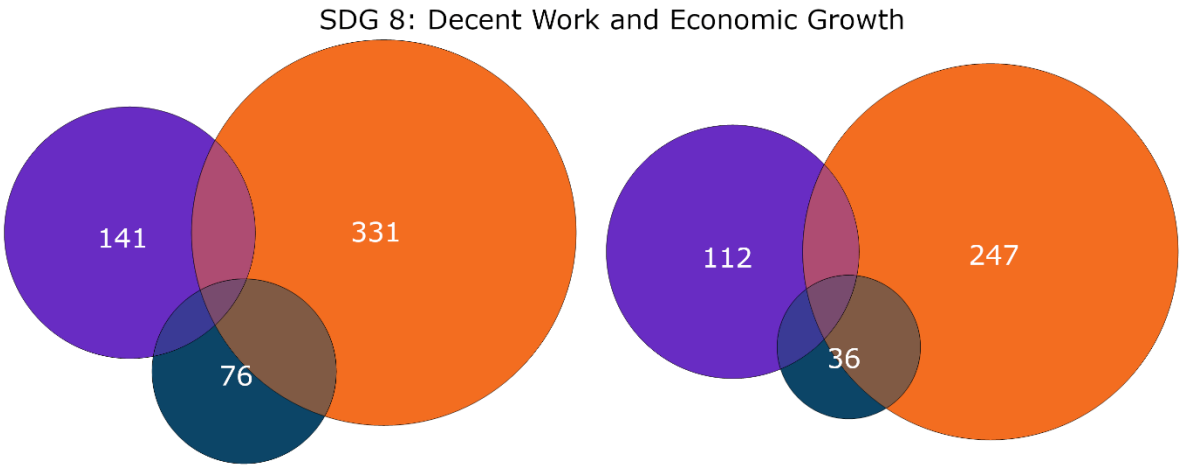

Figure 12. SDG 8: Decent Work and Economic Growth. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 8 has low overlaps with InCites. SciVal has the most SDG-publications. Interestingly, 54% of Dimensions SDG-publications are not covered by the other databases. Dimensions has decent overlap with SciVal, but it also has far fewer SDG-attributed publications.

SDG 9: Industry, Innovation and Infrastructure.

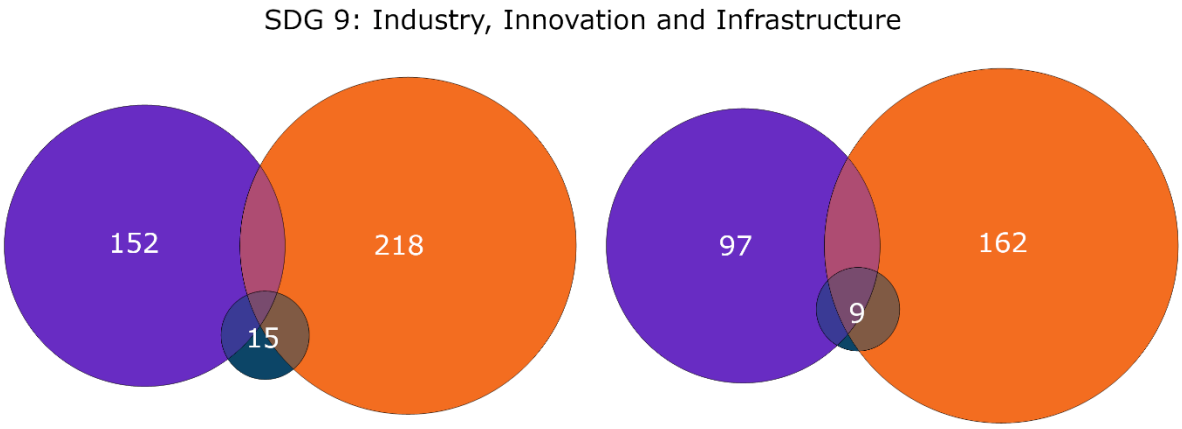

Figure 13. SDG 9: Industry, Innovation and Infrastructure. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 9 is another example of a very low overlap between InCites and Scival (8%/11% and 6%/7%). Dimensions has average overlaps with both InCites and SciVal, but with a very few publications (15/9). Only 64% of InCites SDG-attributed publications are also found in SciVal and Dimensions.

SDG 10: Reduced Inequalities.

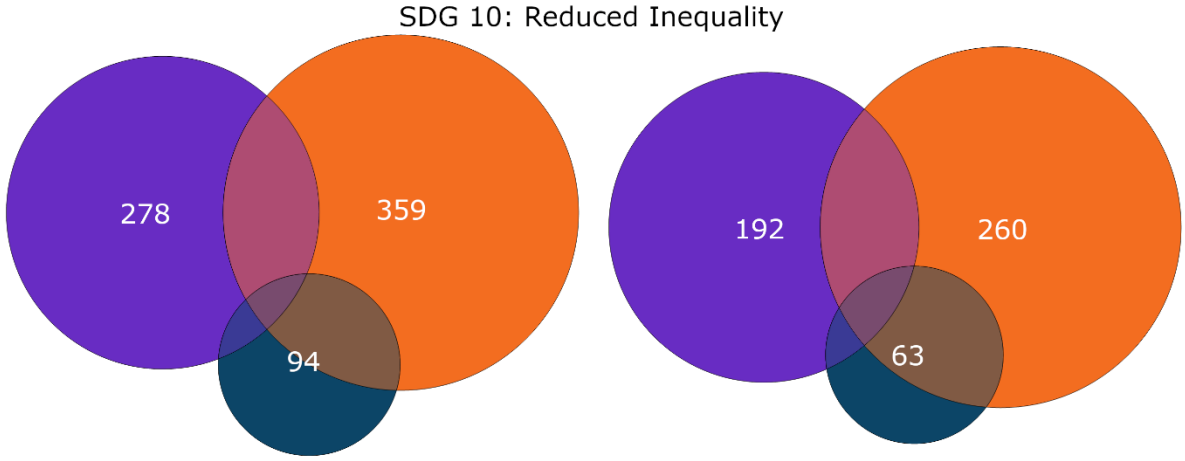

Figure 14. SDG 10: Reduced Inequality. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 10 has rather low overlaps with InCites, but Dimensions and SciVal have average to good overlaps.

SDG 11: Sustainable Cities and Communities.

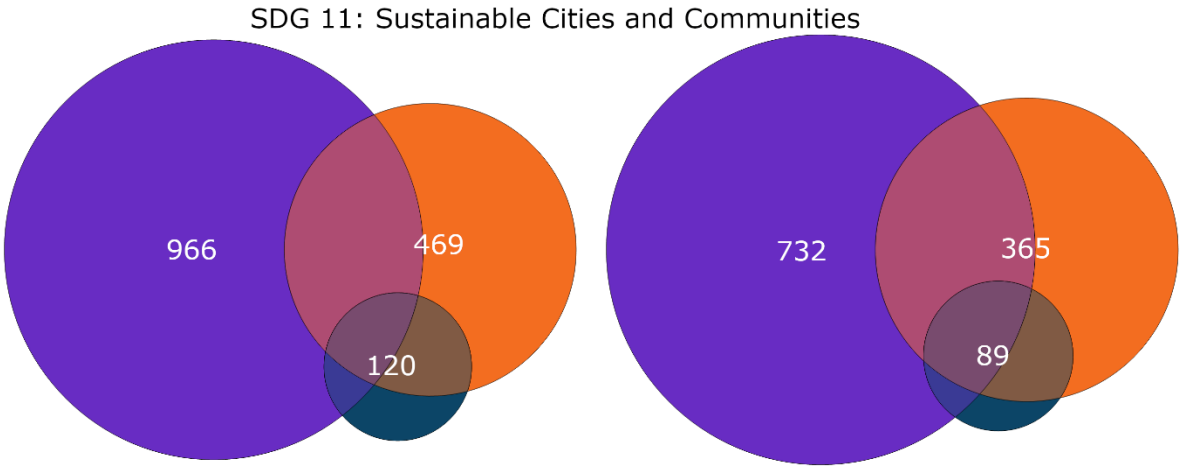

Figure 15. SDG 11: Sustainable Cities and Communities. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 11 has a big difference between InCites and Scival (19%/23% and 40%/46%), but SciVal covers Dimensions rather well (66%/80%). The publications counts are also rather different for the three databases.

SDG 12: Responsible Consumption and Production.

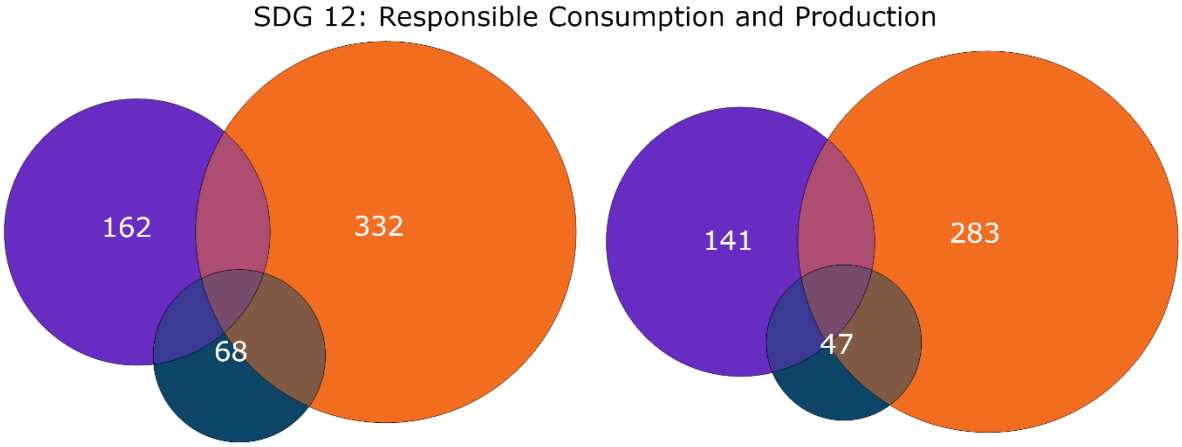

Figure 16. SDG 12: Responsible Consumption and Production. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 12 has very low overlaps between InCites and Scival. Dimensions has better overlaps, but it also has less a lot less publications. Total SDG-publication numbers are also rather different.

SDG 13: Climate Action.

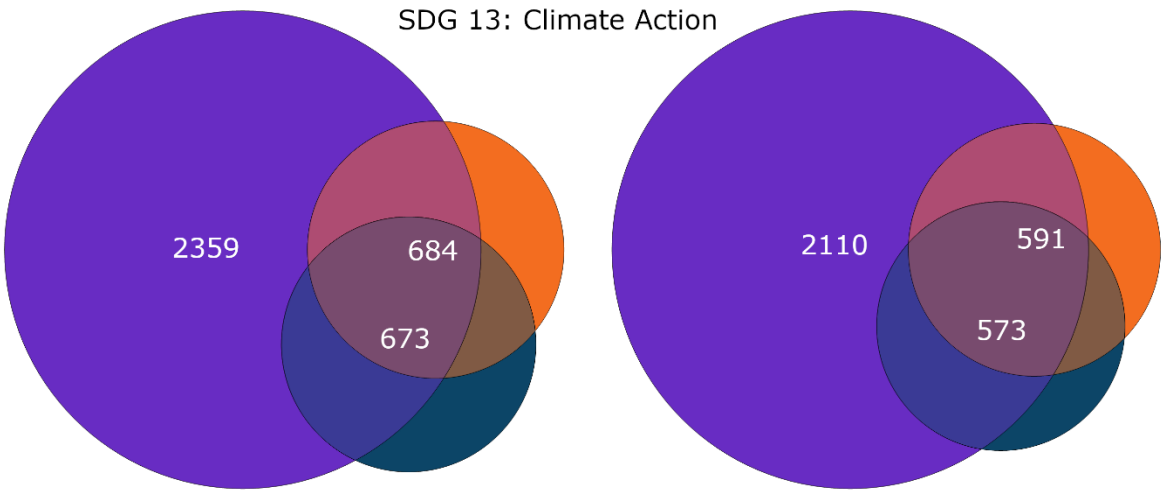

Figure 17. SDG 13: Climate Action. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 13 has rather high overlap between SciVal and Dimensions. They also have very similar publication counts. InCites, however, has a much broader conception of Climate Action. InCites covers SciVal and Dimensions well, but this is probably due to a much larger SDG-publication count.

SDG 14: Life Below Water.

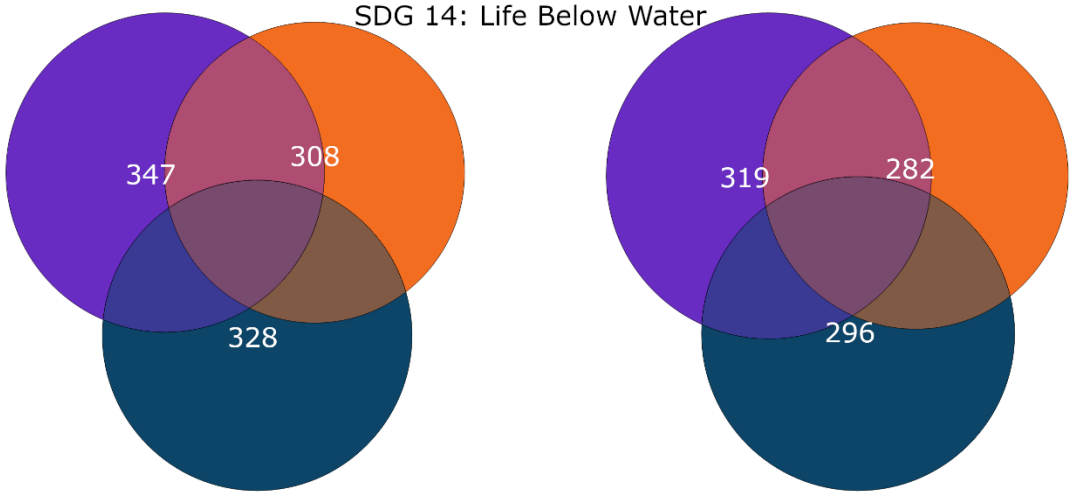

Figure 18. SDG 14: Life Below Water. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 14 has a fairly even spread and overlaps. The publication counts are also very similar. However, the overlaps are at most average (46%).

SDG 15: Life on Land.

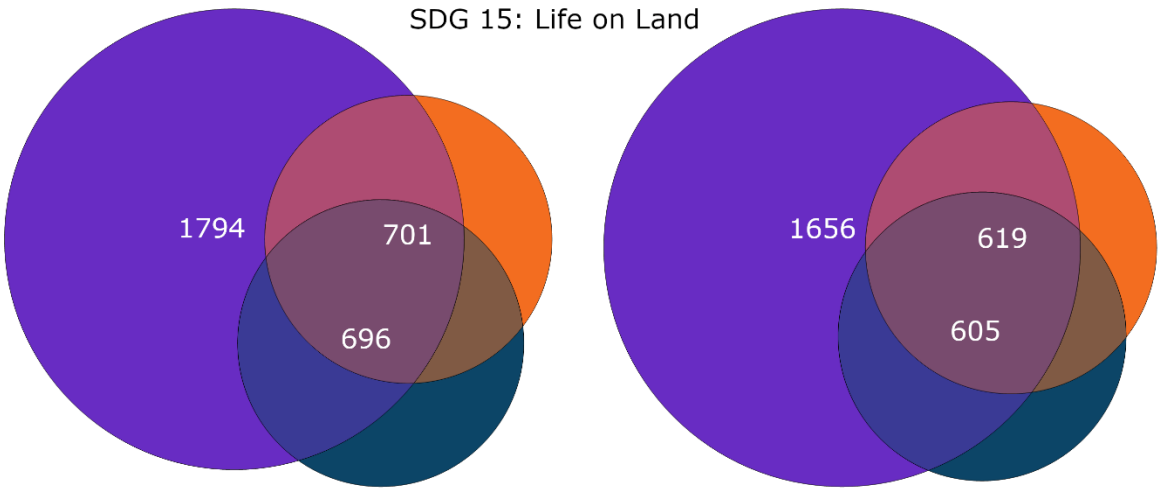

Figure 19. SDG 15: Life on Land. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 15 has fairly good overlaps between Dimensions and SciVal (53%/75% and 54%/59%). InCites has a lot more SDG-publications signalling a wider take on the SDG. Consequently, the overlaps for InCites are low. Publication coverage is rather similar across the databases.

SDG 16: Peace, Justice and Strong Institutions.

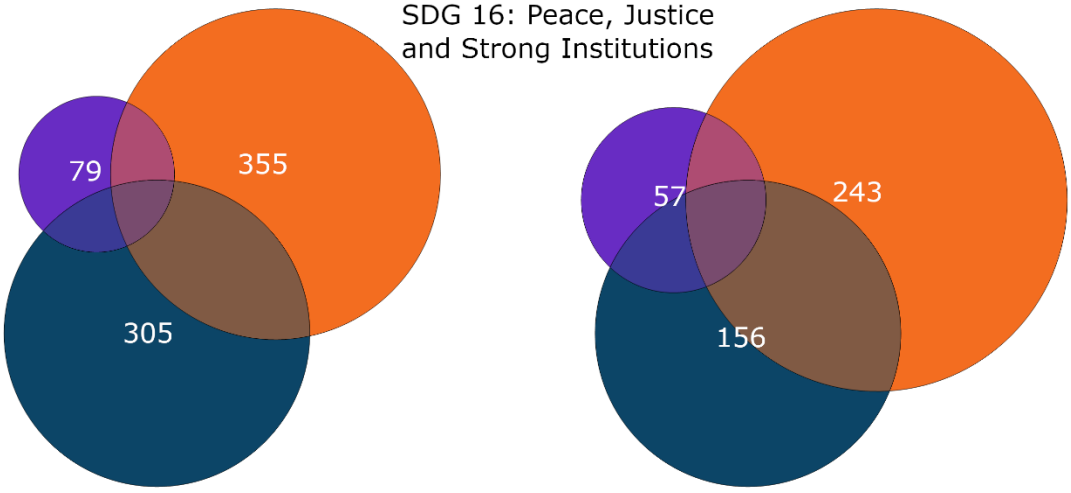

Figure 20. SDG 16: Peace Justice and Strong Institutions. The Venn diagram on the left contains all publications covered by individual databases and the Venn diagram on the right contains only publications that are covered by all three databases. The numbers indicate the amount of SDG-attributed publications in the databases.

SDG 16 is unusual in the sense that unlike most SDGs, InCites has the least amount of SDG attributed publications. InCites SDG-publications also do not overlap here that well with SciVal

(34%/37% and 8%/9%) and Dimensions (32%/44% and 8%/16%)). SciVal and Dimensions also seem to have their own take on the SDG and only have 24%-28% and 28%-44% overlaps.

In general, we can observe that there is a lot of variance at the level of individual SDGs between the different SDG-mapping methods. Overlaps varied but tended to be small to at most average. SDG counts also varied with InCites having the most overall SDG-attributed publications followed by SciVal and Dimensions.

SDG-overlaps are higher when taking only publications that are found in all three databases. However, this might give a misleading impression for a regular use case since SDG-numbers are not typically filtered from multiple databases. Actual variance in SDG-counts between databases also contains difference in coverage.

During the analysis, we also noticed that SDG-attributions can go through rather dramatic changes. Before the InCites 2023 April update, InCites had a total of 7,888 SDG-publications for our university. After the update, that number rose to 15,109 SDG-publications, which is a 91% increase. Similarly, SciVal SDG-attributions have gone through many updates with large effects on total SDG numbers. For Dimensions, we do not have comparative data.

Discussion

It comes as a no surprise that different databases give different SDG-attributions. The databases use different approaches, contain different metadata and have different coverages. There is also an interpretive element in attributing SDGs and interpretation is never an exact activity. As Armitage et al. (2020) has already demonstrated, different SDG-attribution approaches can make a big difference. However, we were still surprised as to the extent of the differences.

Another problem we encounter was that single version updates could have drastic effects on SDG-counts. This meant that the attributions were also highly unstable. This instability might get alleviated in the future once the approaches bed down. There is a reason to be a bit sceptical though. Many of the commonly used classifications in bibliometrics have a more or less direct relation to scientific practice. Classificatory terms such as keywords are shaped by their use and utility in scientific practice. Scientific fields have related organisations, journals and teaching. Research topics and clusters have citation links. SDGs, on the other hand, have no such unifying mechanisms apart from the UN descriptions, which demonstrably admit very differing interpretations. Furthermore, as science progresses and society evolves, our understanding of what research relates to which SDGs is prone to change and this should be reflected in the mappings. The need for interpretation is constant and that makes constancy hard to come by.

It should be pointed out that, in the absence of shared external standards, it is not possible to evaluate which of the SDG-classification schemes are superior (Turunen et al. 2023). If it would so happen that all the approaches would coincide, we could say that the particular choice of a classification scheme is mostly irrelevant. This could have some practical benefits, and possible detriments, but we would still not be able to say that we now have the correct SDG-attributions. This is because we would still lack the external criteria we need for evaluating whether we indeed have the correct attributions. After all, it is not just about getting the bubbles to overlap, but about what we want them to overlap for.

SDG-attributions are supposed to tell us about how research relates to sustainable development, a crucial endeavour, but their source dependence and instability makes them unsuitable for use in indicators and evaluation. At worst, use of SDG-based indicators in evaluation can actively misdirect science policy. This is especially important for the societal impact of science as SDGs are related to sustainable development.

This is not to say that SDG-classification don't have their uses for universities. In fact, there are already a myriad of studies that utilise various SDG-classification for non-evaluative purposes. For example, Aswathy et al. (2023) used keyword co-occurrence mappings between different SDGs and between the topics of a singular SDG, Meschede (2020) looked at author country affiliation co-occurrence clusters for SDG-publications, Payumo et al. (2021) looked at repeated collaboration in SDG-related research and Roy et al. (2022) performed a deep dive into SDG6. Care must be taken to ensure that one is not merely dealing with artefacts due to a choice of a particular classification scheme. However, if we go beyond rigid evaluation, we can take SDG-classifications more as perspectives (see Rafols, et al., 2021) that can highlight important aspects of sustainable development. From this perspective, the differences become a boon.

Acknowledgements

Many thanks to Tuula Huuskonen for her help in gathering relevant data. The cheerful people at NWBRP2024 in Gothenburg also provided invaluable insights for this work. Finally, we want to thank the two anonymous reviewers for their thought-out comments that significantly improved this paper.

About the authors

Petri Turunen is an information specialist at the Helsinki University Library, P.O. Box 53 (Fabianinkatu 30), 00014 University of Helsinki, Finland. He received his Ph.D. from University of Helsinki and his research interests are in metaphilosophy and meta-methodology. He can be contacted at Petri.N.Turunen@Helsinki.fi

Terhi Sandgren is an information specialist at the Helsinki University Library, P.O. Box 53 (Fabianinkatu 30), 00014 University of Helsinki, Finland. She is a doctoral researcher at Tampere University. Her research interests are in information studies. She can be contacted at Terhi.Sandgren@Helsinki.fi

References

Al-Raeei, M. (2023). Analysing of the sustainable development goals in Damascus University during Syrian crisis using the strategy in the university and the bibliometrics data from SciVal. *Discover Sustainability*, 4(24). https://doi.org/10.1007/s43621-023-00140-y

Armitage, C., 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

Aswathy S., Suresh, M., Nedungadi, P., Raman, R. (2023). Mapping analytical hierarchy process research to sustainable development goals: bibliometric and social network analysis. *Heliyon*, 9(8), e19077. https://doi.org/10.1016/j.heliyon.2023.e19077

Bedard-Vallee, A., James, C., Roberge, G. (2023). Elsevier 2023 Sustainable Development Goals (SDGs) mapping. *Elsevier Data Repository*, V1. https://doi.org/10.17632/y2zyy9vwzy.1

Borchardt, S., Barbero Vignola, G., Buscaglia, D., Maroni, M. and Marelli, L. (2022). Mapping EU policies with the 2030 agenda and SDGs. *Publications Office of the European Union*, EUR 31347 EN. Luxembourg. https://doi.org/10.2760/87754, JRC130904.

Confraria, Ciarli, Colonna, Noyons. (2024). The STRINGS queries to identify documents related to the SDGs (+ country-SDG data) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.10470841

Gacia, M. (2022). A more sustainable future for all: introducing the UN Sustainable Development Goals in InCites. Retrieved 27.9.2024 from https://clarivate.com/blog/a-more-sustainable-future-for-all-introducing-the-un-sustainable-development-goals-in-incites/ Internet archive

Hajikhani, A., Suominen, A. (2022). Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. *Scientometrics*, 127(11), 6661-6693. https://doi.org/10.1007/s11192-022-04358-x

Jayabalasingham, B., Boverhof, R., Agnew, K., Klein, L. (2019). Identifying research supporting the United Nations Sustainable Development Goals. *Elsevier Data Repository*, V1, https://doi.org/10.17632/87txkw7khs.1

Ludwig, D., Blok, V., Garnier, M., Macnaghten, P., Pols, A. (2022). What's wrong with global challenges? *Journal of Responsible Innovation*, 9(1), 6-27, https://doi.org/10.1080/23299460.2021.2000130

Meschede, C. (2020). The Sustainable Development Goals in scientific literature: a bibliometric overview at the meta-level. *Sustainability*, 12(11), 4461. https://doi.org/10.3390/su12114461

Nature. (2021). Tracking 20 leading cities' Sustainable Development Goals research. *Nature*, September 24. https://doi.org/10.1038/d41586-021-02406-9

Our World in Data team. (2023). SDG Tracker: Measuring progress towards the Sustainable Development Goals. OurWorldInData.org. Retrieved 27.9.2024 from https://ourworldindata.org/sdgs Internet archive https://web.archive.org/web/20240927101515/https://ourworldindata.org/sdgs

Pakkan, S., Sudhakar, C., Tripathi, S. *et al.* (2023). A correlation study of sustainable development goal (SDG) interactions. *Quality & Quantity*, 57, 1937–1956. https://doi.org/10.1007/s11135-022-01443-4

Payumo, J., He, G., Manjunatha, AC., Higgins, D., Calvert, S. (2021). Mapping collaborations and partnerships in SDG Research. *Frontiers in Research Metrics and Analytics*, 5, 612442. https://doi.org/10.3389/frma.2020.612442

Purnell, P. J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: case study of SDG 13—Climate Action. *Quantitative Science Studies*, 3(4), 976–1002. https://doi.org/10.1162/qss_a_00215

Rafols, I., Noyons, E., Confraria, H., Ciarli, C. (2021). Visualising plural mappings of science for Sustainable Development Goals (sdgs). *SocArXiv*, May 10. https://doi.org/10.31235/osf.io/yfqbd

Raman, R., Nair, V. K., Nedungadi, P. (2023). Discrepancies in mapping Sustainable Development Goal 3 (Good Health and Well-Being) research: a comparative analysis of Scopus and Dimensions databases. *Sustainability*, 15(23), 16413. https://doi.org/10.3390/su152316413

Repiso, R., Segado, F., Gómez-García, S. (2023). La universidad española y la investigación en Web of Science sobre los objetivos de desarrollo sostenible 2017-2021. *Revista Española De Documentación Científica*, 46 (2), e359. https://doi.org/10.3989/redc.2023.2.1980.

Rivest, M., Kashnitsky, Y., Bédard-Vallée, A., Campbell, D., Khayat, P., Labrosse, I., Pinheiro, H., Provençal, S., Roberge, G., James, C. (2021). Improving the Scopus and Aurora queries to identify research that supports the United Nations Sustainable Development Goals (SDGs) 2021. *Elsevier Data Repository*, V4. https://doi.org/10.17632/9sxdykm8s4.4

Roberge, G., Kashnitsky, Y, James, C. (2022). Elsevier 2022 Sustainable Development Goals (SDG) mapping. *Elsevier Data Repository*, V1. https://doi.org/10.17632/6bjy52jkm9.1

Rodenburg, K., De Silva, V., Christensen Hughes, J. (2021). SDGs: a responsible research assessment tool toward impactful business research. *Sustainability*, 13(24), 14019. https://doi.org/10.3390/su132414019

Roy, A., Basu, A., Su, Y., Li, Y., Dong, X. (2022). Understanding recent trends in global Sustainable Development Goal 6 research: scientometric, text mining and an improved framework for future research. *Sustainability*, 14(4), 2208. https://doi.org/10.3390/su14042208

Sachs, J. D. (2012). From millennium development goals to sustainable development goals. *Lancet*, 379(9832), 2206-2211. https://doi.org/10.1016/S0140-6736(12)60685-0

Turunen, P., Hirvonen, I., Pättiniemi, I. (2023). Epistemological Scientism and the Scientific Meta-method. *European Journal for the Philosophy of Science*, 13(2),1-23. https://doi.org/10.1007/s13194-023-00526-9

UN. (2015). Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations: Department of Economic and Social Affairs. Retrieved 27.9.2024 from https://sdgs.un.org/2030agenda Internet archive

Wastl, J., Porter, S., Draux, H., Fane, B., Hook, D. (2020). Contextualizing sustainable development research. *Digital Science*. Report. https://doi.org/10.6084/m9.figshare.12200081.v2

Copyright