

# Annex L: Conference proceedings in Scopus

## Introduction

1. In the REF pilot, we only included articles and reviews, as classified by each of the database providers, in the analysis. At the time we performed the analysis, only Scopus included conference proceedings in their main database. Since then, Thompson-Reuters have integrated their conference proceedings product into Web of Science. In this Annex, we take a brief look at the volume of conference proceedings in Scopus, and the subjects in which they used. We do not look at the bibliometric effects of including proceedings, and anticipate doing further work on this using both databases. As Evidence notes in its report (see Annex H), rather less is known about the use of conference proceedings in bibliometric analysis, compared to articles and reviews, owing to the comparatively recent availability of data.
2. In this annex, we take a brief look at the percentage of outputs in the Scopus database as a whole that are classified as conference proceedings. We then focus on Engineering and Computer Science, where we know conference proceedings are more likely to play a significant role in research dissemination practices. We contrast this with work published in Physics, where we believe that this is less likely to be the case.
3. We also consider the specific case of conference proceedings being published in volumes classified as journals. This may occur when, for example, a journal publishes a special 'conference' edition. The aim of this part of the work is to help inform how we handle such outputs for bibliometric purposes: do they behave as normal articles, or as conference proceedings? We again focus on the subjects listed above, to compare publication practices in a range of subject areas.
4. This work highlights some of the potential difficulties in handling conference proceedings. It is not clear whether conference proceedings published in normal journals should be treated as proceedings or as articles

## Scopus database fields and overview for this piece of analysis

### Scopus source and document types

5. Scopus has seven publication source types and 20 document types. These are listed in Table L1. We first look at the entire Scopus database, without reference to source type.

**Table L1: Scopus source and document types**

Source types		Document types	
1	Book	1	Abstract Report
2	Trade Journal	2	Article
3	Journal	3	Book
4	Book Series	4	Book Review
5	Multi-volume Reference Works	5	Business Article
6	Conference Proceedings	6	Chapter

7	Report	7	Conference Paper
		8	Conference Review
		9	Dissertation
		10	Editorial
		11	Erratum
		12	Article in Press
		13	Letter
		14	Note
		15	Patent
		16	Press Release
		17	Review
		18	Report
		19	Short Survey

6. In order to look at disciplinary differences, we use the ASJC codes associated with each output in order to extract records relating to Engineering, Computer Science and Physics.

7. Table L2 shows the ASJC codes that we took to represent each subject area. Each output is associated with one or more ASJC codes. These are determined by the publication source (for example journal or conference proceeding) in which it is published. Outputs that have been assigned to several ASJC codes are mapped to our subject areas if any of the ASJC codes linked to them are included in our lists. Such outputs are, however, only included once in the subject area, even if several of their ASJC codes map to the subject area.

**Table L2: Selected ASJC codes**

ASJC Code	Engineering	ASJC Code	Computer Science	ASJC Code	Physics
2200	Engineering (all)	1700	Computer Science (all)	3100	Physics and Astronomy (all)
2201	Engineering (miscellaneous)	1701	Computer Science (miscellaneous)	3101	Physics and Astronomy (miscellaneous)
2202	Aerospace Engineering	1702	Artificial Intelligence	3102	Acoustics and Ultrasonics
2203	Automotive Engineering	1703	Computational Theory and Mathematics	3103	Astronomy and AstroPhysics

2204	Biomedical Engineering	1704	Computer Graphics and Computer-Aided Design	3104	Condensed Matter Physics
2205	Civil and Structural Engineering	1705	Computer Networks and Communications	3105	Instrumentation
2206	Computational Mechanics	1706	Computer Science Applications	3106	Nuclear and High Energy Physics
2207	Control and Systems Engineering	1707	Computer Vision and Pattern Recognition	3107	Atomic and Molecular Physics, and Optics
2208	Electrical and Electronic Engineering	1708	Hardware and Architecture	3108	Radiation
2209	Industrial and Manufacturing Engineering	1709	Human-Computer Interaction	3109	Statistical and Nonlinear Physics
2210	Mechanical Engineering	1710	Information Systems	3110	Surfaces and Interfaces
2211	Mechanics of Materials	1711	Signal Processing		
2212	Ocean Engineering	1712	Software		

8. Table L3 shows the number and percentage of outputs in our snapshot of Scopus that are mapped to each document type, and the number and percentage of outputs mapped to each document type in our three exemplar subject areas.

**Table L3: Overall number and percentage of document types in all source types found in Scopus; for the database as a whole and specific fields**

Document Type	All	Eng %	Comp Sci	Physics	All %	Eng %	Comp Sci %	Physics %
Abstract								
Report	1,299	651	#N/A	#N/A	0%	0%	#N/A	#N/A
Article	8,017,408	899,574	301,746	931,327	68%	47%	49%	72%
Book	592	32	#N/A	#N/A	0%	0%	#N/A	#N/A
Business Article	27,933	2,255	#N/A	1	0%	0%	#N/A	0%
<b>Conference Paper</b>	<b>1,631,007</b>	<b>822,866</b>	<b>261,883</b>	<b>313,539</b>	<b>14%</b>	<b>43%</b>	<b>42%</b>	<b>24%</b>

Conference Review	14,674	9,415	3,436	3,874	0%	0%	1%	0%
Editorial	233,849	22,807	11,535	6,901	2%	1%	2%	1%
Erratum	66,868	2,836	1,193	7,265	1%	0%	0%	1%
Article in Press	6,762	802	539	388	0%	0%	0%	0%
Letter	263,326	9,961	3,376	3,856	2%	1%	1%	0%
Note	345,782	48,133	7,477	4,879	3%	3%	1%	0%
Patent	799	#N/A	#N/A	#N/A	0%	#N/A	#N/A	#N/A
Press Release	7,403	#N/A	#N/A	#N/A	0%	#N/A	#N/A	#N/A
Review	1,047,084	75,647	23,628	24,644	9%	4%	4%	2%
Report	47	3	#N/A	#N/A	0%	0%	#N/A	#N/A
Short Survey	196,020	23,921	4,581	2,854	2%	1%	1%	0%
<b>Totals</b>	<b>11,860,853</b>	<b>1,918,903</b>	<b>619,394</b>	<b>1,299,528</b>	100%	100%	100%	100%

9. We see that a large percentage of outputs (more than 40 per cent) in both Engineering and in Computer Science are conference proceedings. A smaller percentage (24 per cent) of outputs in Physics is classified as such; however, this is still a larger percentage than for the database as a whole (14 per cent).

### Conference proceedings in normal journals

10. Having looked at the percentage of work assigned to each document type in the database as a whole, we turn our attention to the specific case of document types published in 'normal' journals. We do this by looking only at items of source type 'journal'. We also only wish to consider journals that are published regularly and frequently throughout the census period, in order to focus on journals that are likely to be a core part of the literature.

### Journal coverage

11. The range and frequency of journal coverage will vary in Scopus. If journals were not covered fully or only published infrequently through the time period of 2001-2007, they were excluded from this analysis. By only including journals that are fully covered, we get a core data set that we believe to be suitable for this analysis. Journals were identified by their ISSN.

12. Since this study looks at the numbers and percentages of document types in journals, any outputs with missing information vital to calculation were also excluded. For instance, outputs with missing data in the volume, issue or publication year fields were not included.

## Method details

13. To find conference papers in well-covered journals, we extracted a list of journals with outputs published within 2001-2007. At this stage the data also included details on the document type (for example article or editorial), volume, issue, and year of publication of when the item was published. Calculations were then performed to give summary information by journal, volume, issue, and publication year to show how the number of documents of each type and their percentages made up an issue of that journal. For example:

**Table L4: Example output of document type number and percentage per journal issue**

Source title	Journal (ISSN)	Publication year	Volume	Issue	Document type	Number	Total	%
XXXX	YYYY	2003	10	1	Article	5	10	50
XXXX	YYYY	2003	10	1	Review	3	10	30
XXXX	YYYY	2003	10	1	Conference Paper	2	10	20

14. Once the calculations were performed, the data set was then searched to identify a well-covered, frequently publishing, pool of journals over the 2001-2007 time period. We aimed to exclude journals publishing only annual issues, not covered thoroughly or those with unusually high number of issues due to erroneous subdivision of issues (for example issue 1 I, issue 1 II). The well-covered journals were identified on the basis of how many issues they produced during the time period. The journal was included if the number of issues spanning the seven-year time period was 28-450 issues (inclusive). A journal having 28 issues would indicate the journal was published, on average, quarterly over the time period. A journal having 450 issues would indicate it was published weekly and had at most one additional issue or supplement per month.

15. The assumption is that a special issue or supplement would not occur more than once per month for all months of the year. Only three journals had over 450 issues in the time period. By focusing on these journals, we reduce the number of outputs and journals in our data set as shown in Table L5. The percentage of outputs by document type in these journals is shown in Table L6.

16. We then focused on the exemplar subject areas we used when looking at the database as a whole: Engineering, Computer Science and Physics. We used the same mapping of ASJCs to look at outputs published in journals mapped to these subject areas, and applied the same methodology to these data sets. The results for these subject areas are in Table L7.

**Table L5: Overall number of outputs found and number of distinct journals included in the study**

<b>All outputs</b>	7,512,216
<b>Distinct journals</b>	8,250

**Table L6: Number and percentage of outputs by document type for well-covered, frequently publishing journals within 2001-2007**

<b>Document types</b>	<b>Number of outputs</b>	<b>%</b>
Abstract Report	211	0
Article	5,766,524	77
Book	17	0
Business Article	811	0
<b>Conference Paper</b>	<b>365,969</b>	<b>5</b>
Conference Review	497	0
Editorial	155,257	2
Erratum	56,263	1
Letter	216,458	3
Note	179,781	2
Review	655,361	9
Short Survey	115,067	2
<b>Total</b>	<b>7,512,216</b>	

**Table L7: Number and percentage of document types in well-covered journals found in Scopus for Engineering, Computer Science and Physics, within 2001-2007**

<b>Document Type</b>	<b>Eng</b>	<b>Comp Sci</b>	<b>Physics</b>	<b>Eng %</b>	<b>Comp Sci %</b>	<b>Physics %</b>
Article	558,612	190,194	754,461	85%	82%	84%
Business Article	10	NA	1	0%	NA	0%
<b>Conference Paper</b>	<b>41,497</b>	<b>17,828</b>	<b>102,992</b>	<b>6%</b>	<b>8%</b>	<b>12%</b>
Conference Review	114	108	231	0%	0%	0%
Editorial	9,000	5,630	4,361	1%	2%	0%
Erratum	2,268	899	6,672	0%	0%	1%
Letter	5,142	1,880	3,212	1%	1%	0%

Note	11,739	2,062	3,173	2%	1%	0%
Review	26,632	11,521	18,067	4%	5%	2%
Short						
Survey	4,976	906	1,716	1%	0%	0%
<b>Totals</b>	<b>659,990</b>	<b>231,028</b>	<b>894,886</b>	100%	100%	100%

## Discussion

17. As noted in paragraph 10, for this part of the analysis where we focus on the specific case of conference papers being published in normal, well-covered journals, we limited the sources we included to those which appeared to be published regularly in order to focus on journals that are more likely to be a core part of the subject literature. A summary of the overall number of outputs and number of distinct journals for all subjects is given in Table L5.

18. For all subjects, as we can see from Table L6, there are outputs identified as being conference papers; however, they do not make up a large percentage of the overall items.

19. If we now focus on the exemplar subjects, the percentage of conference papers in 'normal' journals for Physics is larger than the other exemplar subject areas (Table L7). This is in contrast with Scopus' snapshot of all outputs (Table L3) where a smaller percentage of outputs in Physics were classified as conference proceedings than the other exemplar subject areas. This suggests that conference proceedings in 'normal' journals are a more common phenomenon in Physics than in Computer Science and Engineering, where conference papers appear to be more likely to be published in volumes of conference proceedings.

20. It must be remembered that all of the work in this annex is dependent on Scopus' classification of publication source type and document type.

21. This annex gives an overview of the number and percentages of conference papers appearing in the database as a whole. This indicates that conference proceedings make up a significant percentage of the database as a whole, which varies between subject areas.

22. If we look at the specific case of conference papers being published in normal journals, we see that the percentage of conference papers in the exemplar subject areas we looked at changes significantly. This suggests that there is a subject-specific dimension to the publication sources used to disseminate conference proceedings.

23. We plan to look at whether these outputs behave, from a bibliometrics perspective, as regular research articles, or as conference proceedings. This is likely to inform how we normalise such work.

24. This initial study will inform future work we will undertake, where we will look in more detail at the use of conference proceedings in different disciplines, and at how they behave from a bibliometric analysis perspective.

25. It is important to remember that this annex has only looked at Scopus' coverage of conference proceedings. In an exercise such as the REF, where we propose to focus on researchers' best outputs, it may well be the case that the percentage of conference

proceedings submitted in each subject area is rather different from bibliometric databases' coverage of such outputs.