

# Comparing Coverage of Scopus, WoS, and OBRSS List: A Case for Institutional and National Databases of Research Output?

Lai Ma<sup>1</sup> and Liam Cleere<sup>2</sup>

<sup>1</sup>*lai.ma@ucd.ie*

School of Information and Communication Studies, University College Dublin, Dublin 4, Ireland

<sup>2</sup>*liam.cleere@ucd.ie*

Office of the Vice President for Research and Innovation, University College Dublin, Dublin 4, Ireland

## Abstract

University College Dublin (UCD) has implemented the Output-Based Research Support Scheme (OBRSS) since 2016. Adapted from the Norwegian model, the OBRSS aims to incentivise research and publications by disbursing research support funds based on research outputs including publications and PhD supervision. This paper examines the coverage of UCD publications in Scopus, Web of Science (WoS), and the OBRSS list. The analysis shows that the OBRSS list has a more comprehensive coverage in all disciplines and has a significant advantage in the coverage of SSH disciplines compared to Scopus and WoS. It is also evident that the development of the OBRSS list is more transparent about its indexing practices and procedures. Further, the list offers opportunities for academic and research community to co-construct the list and to justify the prestige of publications, which allows the inclusion of novel, non-commercial, and open access publications without track records of citations and journal impact factor. The OBRSS has also encourages regular updates of publication records by academic and research staff, which leads to mostly complete information about publication activities useful for bibliometric analysis, research management, and strategic planning.

## Introduction

The Output-Based Research Support Scheme (OBRSS) has been implemented in University College Dublin (UCD) since 2016. Adapted from the Norwegian model, the OBRSS aims to incentivise research and publication by disbursing research support funds based on research outputs including publications and PhD supervision. The design of the OBRSS involved the construction of a ranked publication list and a points system, and its implementation is contingent on regular and reliable updates on the Current Research Information System by academic and research staff.

At the time of this writing, there is not a national initiative to collate complete data of academic and research publications in Ireland, nor is there a systematic effort to create a database such as CRISTin in Norway (Sivertsen, 2018) or the Publication Forum in Finland (Pölonen, 2018). Commercial databases and tools such as Elsevier's Scopus and SciVal and Thomas Reuter's Web of Science (WoS) are used as data sources for counting research outputs and depicting publication trends, despite the common knowledge that the coverage of these commercial databases is incomplete, especially for disciplines in the social sciences and the humanities (Mongeon & Paul-Hus, 2016; Sivertsen & Larsen, 2012). Zacharewicz, et. al (2018) has shown that over 70% of research funding in Ireland has been channelled to project funding, which largely benefits STEM research. While there is a strong emphasis on STEM-oriented projects, many disciplines in the social sciences and humanities in Ireland are ranked higher in subject rankings internationally. There is, however, a lack of data about publication patterns and other research outputs about all disciplines.

Based on a European survey, Sile, et. al (2018) show that not all EU member states maintain national bibliographic databases and the completeness and uses of existing ones vary. Bibliographic data can be linked to data reporting, research evaluation, and research funding

allocation in some countries but not others. It has also been noted that the COST Action ENRESSH (European Network for Research Evaluation in the Social Sciences and Humanities) envisions a European database by integrating existing databases and information systems in Europe (p. 12). The VIRTU-ENRESSH proof of concept pilot (Puuska, et al, 2018) and the Nordic List (NSD, 2018) are pioneering projects in this area.

What are the benefits of institutional/national bibliographic databases? Sivertsen (2010) has discussed the need for complete data when designing performance indicator. A few case studies have also been conducted to understand the validity of bibliometric analyses in local information systems such as METIS in the Netherlands (van Leeuwen, van Wijk & Wouters, 2016). There is a general agreement that national/institutional databases or information systems are necessary for bibliometric analysis in the social sciences and the humanities (see, for example, Sivertsen & Larsen, 2012; Ossenblok, Engels & Sivertsen, 2012) because the coverage of commercial databases is less than satisfactory in some disciplines, not to mention the indexing practices of commercial providers can be driven by market interests.

The uses of national/institutional databases are relatively under-explored in bibliometrics. This paper examines the coverage of UCD publications in Scopus, Web of Science, and the OBRSS list, followed by a discussion of the potential benefits of institutional and national databases for research evaluation and open science, as well as related issues pertaining to fairness and transparency of research management and research policy.

#### **Background: OBRSS**

University College Dublin (UCD) recognises that faculty’s commitment to excellent research helps build a strong research reputation. The university also recognises that many of the day-to-day costs of research activity are not covered by research grants. In recognition of this, UCD has developed the Output-Based Research Support Scheme (OBRSS) to disburse research support funds to faculty based on their research outputs, as captured through publications and PhD supervision.

Publications are divided into two categories: normal and prestigious. The OBRSS uses lists of publication channels – one for Publishers and one for Series (Journals, Book Series, and Conference Series) – as a reference for the categorisation of publications. The lists are dynamic and are updated once every year by using the Danish, Finnish & Norwegian peer reviewed publication channel lists as a baseline. Refinements are made to this by UCD faculty each year.

The OBRSS uses the ranked publication list – one section for Publishers and another for Series (Journals, Book Series, and Conference Series) – as a reference for the calculation of points. Each publication is assigned one of two levels: level 1 – Normal or level 2 – Prestigious. Weighted scores are then applied to each publication. Similar to the Norwegian model, points are allocated for different types of publication as summarised in Table 1:

**Table 1 Points allocation per publication type**

Publication types	Points Level 1 ‘normal’	Points Level 2 ‘prestigious’
<b>Book</b>	5	8
<b>Journals Article</b>	1	3
<b>Book Chapter</b>	1	3
<b>Conference Publication</b>	0.5	2
<b>Edited Book</b>	1	3
<b>Other Publication</b>	0.5	2
<b>Published Report</b>	1	3

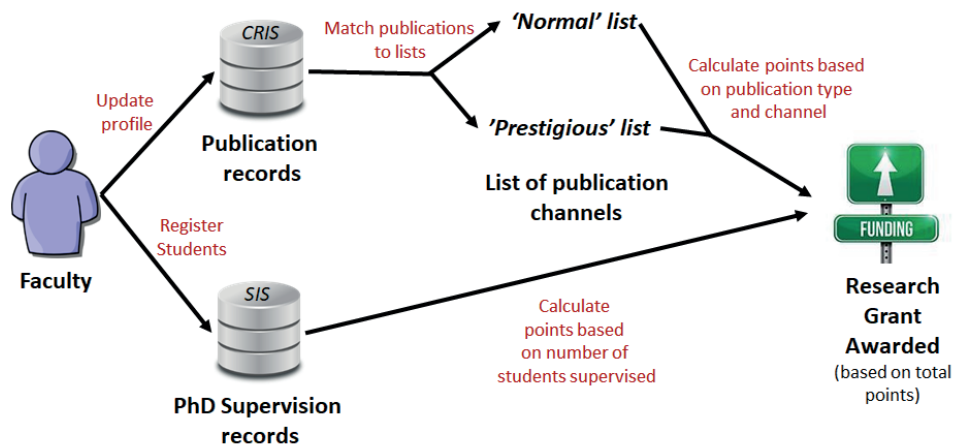


Figure 1 OBRSS Points

There is a consultation process to ensure that inputs from the academic staff are considered in finalising the ranked publication list. During the consultation period, academic staff can make recommendations to add/remove publications to/from the ranked publication list at the two levels. The suggestions and recommendations are reviewed by the Office of the Vice President for Research. Considering the objectives and scope of the OBRSS, external panels are not used to review the ranked publication list. Overall, approximately 1% of the total annual research budget for the university is allocated to the OBRSS.

Table 2 Calculation of publication output point

*Publication output-points = B x C x F x N, where*

- *B = Points (allocated based on the type of publications and whether it is in a 'normal' or 'prestigious' channel)*
- *C = collaboration factor (multiply by 1.25 if there are any international authors on the paper)*
- *F = UCD author factor (multiply by 0.7 if there are two UCD academic staff on the paper; multiply by 0.6 if there are three UCD academic staff on the paper; multiply by 0.5 if there are four or more UCD academic staff on the paper)*
- *N = if the total number of authors on a paper exceeds 100, multiply the result by 0.1*

Publication points are calculated for each academic staff's publications in the CRIS over a three-year period (for example 2015-2017) using the formula in Table 2. All academic staff are automatically entered into the OBRSS each year. The total points that an academic staff has accumulated is communicated using a personalised points statement. Final points statements are issued to academic staff receiving an award in October each year. The minimum value threshold for a research award is €200 and there is no maximum research award.

### **Coverage of Scopus, Web of Science and OBRSS List of UCD Research Outputs**

One of the most significant outcomes of the implementation of the OBRSS is a more complete picture of publication records in University College Dublin. The number of academic staff updating their research profiles in the CRIS has increased each year. In the first year the OBRSS was implemented, 85% of academic staff updated their profiles as opposed to 75% over the previous three years. The publication records are essential to understand publication practices, in terms of publication types and frequency, for example, in different disciplines. While using the Danish, Finnish, and Norwegian lists as the baseline, the OBRSS list has been updated regularly with new publications recorded in CRIS, as well as inputs from academic staff.

The OBRSS ranked publications list covers over 78.6% of all UCD publications, with the highest at 88.7% in the College of Science and the lowest at 67.5% in the College of Business (Table 3). The OBRSS list is primarily designed to traditional publication outputs such as books and journal publications. Its coverage of journal articles is 97.9% while coverage of book-based publications average 80%. Since STEM disciplines publish more of their content on average in journals, this tends to improve the overall coverage for these disciplines.

Publications not 'counted' by the OBRSS list include 45% of conference papers; 80% of 'other publications' and 75% of published reports (Table 4). The lack of coverage is due to not having recognised publishers or unique identifiers (e.g. ISSNs) associated with the publications.

While the coverage of Scopus and WoS of UCD publications varies, with biggest differences in the College of Engineering and Architecture and a small difference in the College of Arts and Humanities, the rankings of coverage is the same, from highest to lowest coverage: Science, Health and Agricultural Sciences, Engineering and Architecture, Business, Social Sciences and Law, and Arts and Humanities.

The rankings by School (Table 5) give a more nuanced comparison of the coverage of Scopus and WoS. Publications in Computer Science, for example, are covered much higher in Scopus than WoS. The top ten subjects in both Scopus and WoS databases are STEM-related, with the exception of Economics in Scopus. It is clear that the coverage of arts and humanities subjects is very weak, as has been surveyed by many (see Mongeon & Paul-Hus, 2016 for a recent analysis). This is in stark contrast to the OBRSS coverage where there is much higher coverage in all disciplines including arts and humanities.

In summary, the analysis shows that there has been a lack of data sources for understanding publication trends in the social sciences and the humanities, at least in the context of UCD. The coverage and discrepancies between the two major commercial databases are not negligible. It is also clear that research evaluation at individual and institutional level (e.g., university rankings) should not depend solely on these databases, particularly taking into account the levels of coverage in each subject, not to mention the differences in citation practices that would affect h-index, for example. Overall, the OBRSS list represents a higher percentage of publications in all subjects, but has a significant advantage in the social sciences and the arts and humanities.

**Table 3 Comparison of publication volumes per college for academic staff only from 2015 to 2017 inclusive; using CRIS/OBRSS data from UCD RMS Profiles June 2018, Scopus data from June 2018; and Web of Science data (WoS) from Jan 2019**

UCD College Name	CRIS Total 2015-17	OBRSS Total 2015-17	Scopus Total 2015-17	WoS Total 2015-17	% Coverage in OBRSS	% Coverage in Scopus	% Coverage in WoS
Science	3,221	2,856	2,365	1,990	88.7%	73.4%	61.8%
Health and Agricultural Sciences	4,294	3,517	2,489	2,351	81.9%	58.0%	54.8%
Engineering and Architecture	2,469	1,834	1,545	1,048	74.3%	62.6%	42.4%
Arts and Humanities	893	634	134	147	71.0%	15.0%	16.5%
Social Sciences and Law	2,168	1,489	750	548	68.7%	34.6%	25.3%
Business	652	440	278	210	67.5%	42.6%	32.2%
<b>Grand Total</b>	<b>13,697</b>	<b>10,770</b>	<b>7,561</b>	<b>6,294</b>	<b>78.6%</b>	<b>55.2%</b>	<b>46.0%</b>

**Table 4 Comparison of publication volumes per types for academic staff only from 2015 to 2017 inclusive; using CRIS/OBRSS data from UCD RMS Profiles June 2018, Scopus data from June 2018; and Web of Science data (WoS) from Jan 2019**

UCD School Name	CRIS Total 2015-17	OBRSS Total 2015-17	Scopus Total 2015-17	WoS Total 2015-17	% Coverage in OBRSS	% Coverage in Scopus	% Coverage in WoS
Book	207	140	53	9	67.6%	25.6%	4.3%
Chapter	1,273	1,027	292	135	80.7%	22.9%	10.6%
Conference Paper	3,506	1,931	891	659	55.1%	25.4%	18.8%
Edited Book	209	184	28	5	88.0%	13.4%	2.4%
Journal article	7,541	7,308	6,296	5,469	96.9%	83.5%	72.5%
Other	725	144	1	17	19.9%	0.1%	2.3%
Report	236	36			15.3%	0.0%	0.0%
<b>Grand Total</b>	<b>13697</b>	<b>10770</b>	<b>7561</b>	<b>6294</b>	<b>78.6%</b>	<b>55.2%</b>	<b>46.0%</b>

## Discussion and Conclusion

The OBRSS has the objective of incentivising publications in high-quality, international publication outlets; at the same time, it also encourages regular updates of publication records by academic and research staff, which leads to close to complete information about publication records within the university. Before the implementation of the OBRSS, data were incomplete as academic staff were less motivated to keep their research profiles up-to-date.

**Table 5 Comparison of publication volume per school for academic staff 2015-2017 inclusive: using CRIS/OBRSS data from UCD RMS Profiles June 2018, Scopus data from June 2018; and Web of Science data (WoS) from Jan 2019**

UCD School Name	CRIS Total 2015-17	OBRSS Total 2015-17	Scopus Total 2015-17	WoS Total 2015-17	% Coverage in OBRSS	% Coverage in Scopus	% Coverage in WoS
Physics	883	854	810	733	96.7%	91.7%	83.0%
Chemistry	272	254	229	236	93.4%	84.2%	86.8%
Biomolecular and Biomedical Science	434	396	309	295	91.2%	71.2%	68.0%
Medicine	1,621	1,471	1,066	926	90.7%	65.8%	57.1%
Biology and Environmental Science	304	275	205	183	90.5%	67.4%	60.2%
Economics	114	102	80	66	89.5%	70.2%	57.9%
Electrical and Electronic Engineering	574	507	489	300	88.3%	85.2%	52.3%
Music	44	38	6	5	86.4%	13.6%	11.4%
Philosophy	224	190	54	37	84.8%	24.1%	16.5%
Earth Sciences	283	239	84	91	84.5%	29.7%	32.2%
Chemical and Bioprocess Engineering	163	134	130	107	82.2%	79.8%	65.6%
Veterinary Medicine	736	603	406	432	81.9%	55.2%	58.7%
Mathematics and Statistics	426	349	291	204	81.9%	68.3%	47.9%
Biosystems and Food Engineering	415	334	278	228	80.5%	67.0%	54.9%
Computer Science	619	489	437	248	79.0%	70.6%	40.1%
Public Health, Physiotherapy and Sports Science	676	534	354	380	79.0%	52.4%	56.2%
English, Drama and Film	252	192	55	59	76.2%	21.8%	23.4%
Sociology	138	104	48	45	75.4%	34.8%	32.6%
Politics and International Relations	182	137	77	63	75.3%	42.3%	34.6%
Languages, Cultures and Linguistics	210	157	37	36	74.8%	17.6%	17.1%
Archaeology	118	88	23	13	74.6%	19.5%	11.0%
Agriculture and Food Science	945	688	537	504	72.8%	56.8%	53.3%
Mechanical and Materials Engineering	609	440	348	195	72.2%	57.1%	32.0%
History	200	142	28	39	71.0%	14.0%	19.5%
Nursing, Midwifery and Health Systems	316	221	126	109	69.9%	39.9%	34.5%
Psychology	389	267	176	157	68.6%	45.2%	40.4%
Civil Engineering	351	238	194	132	67.8%	55.3%	37.6%
Art History and Cultural Policy	77	52	3	7	67.5%	3.9%	9.1%
Business	652	440	278	210	67.5%	42.6%	32.2%
Education	100	67	37	23	67.0%	37.0%	23.0%

Classics	24	16	4		66.7%	16.7%	0.0%
Information and Communication Studies	98	64	48	20	65.3%	49.0%	20.4%
Law	306	182	35	12	59.5%	11.4%	3.9%
Geography	230	133	95	80	57.8%	41.3%	34.8%
Social Policy, Social Work and Social Justice	269	155	77	32	57.6%	28.6%	11.9%
Architecture, Planning and Environmental Policy	357	181	106	86	50.7%	29.7%	24.1%
Irish, Celtic Studies and Folklore	86	37	1	1	43.0%	1.2%	1.2%
<b>Grand Total</b>	<b>13,697</b>	<b>10,770</b>	<b>7,561</b>	<b>6,294</b>	<b>78.6%</b>	<b>55.2%</b>	<b>46.0%</b>

The OBRSS ranked publication list also gives a better overview of research activities of all disciplines, compared to data generated from commercial providers such as Scopus or WoS. This study shows that the coverage of the OBRSS list is significantly more comprehensive in SSH disciplines. These data can lead to better strategic planning, research management and research policy. For example, UCD School of History is ranked in the top 100 in the QS 2019 subject rankings, but only 14% of its publications were indexed by Scopus and 19.5% indexed by WoS. The data generated by the institutional database would be more useful to analyse the publication trends and other factors that contribute to high academic reputation, for example.

National and institutional databases with comprehensive metadata can also allow for analyses based on factors such as gender and career stage. Currently, most commercial databases and services such as ORCID do not record gender, career stage, and other useful factors for bibliometric analyses. The lack of these data makes analyses complicated and prone to errors.

While an ideal list of publications is difficult to attain, the OBRSS list, like the Norwegian and Finnish list, offers opportunities for academics and researchers to make suggestions and comments. Hence, the prestige of a journal does not solely depend on the journal impact factor, or it being indexed by commercial databases; rather, to a certain extent, the prestige of publications is a consensus of the academic and research community. Ideally, disagreement and discontentment are resolved in open discussion with the goal of maintaining a fair and representative list of publications.

The co-construction of an institutional and national list also allows the inclusion of new publications that have not accumulated citations and hence have not been indexed by commercial databases. These could include publishers who are highly recommended by experts, and those who are enthusiastic about open science and open access. The inclusion of new publishers, apart from the Big Deals, would be beneficial to knowledge production as it allows new voices to be heard—and be rewarded.

Currently, research profiles involving individual and institutional research performances are usually generated by commercial services such as Scopus and Google Scholar. However, their indexing practices are not transparent compared to national or institutional databases, for example, in the case of the Norwegian list (Sivertsen, 2010) and, for instance, the OBRSS list. National and institutional databases could be a trusted source with fair and transparent procedures with inputs from the academic and research community.

To conclude, the construction of the OBRSS list would not be possible without sufficient resources and support. The implementation of the OBRSS encourages regular updates on publication records, which also assists refinement of the ranked publications list. However, like other performance-based funding systems, the effects on publication trends and research practices would require examination over time (see, for example, Aagaard, 2015; Butler, 2003; Hammerfelt & de Rijcke, 2015; Hicks, 2012; Ma, 2018). Nevertheless, the comparison of coverage shows that there are benefits of institutional and national databases. First, the publications represent institutional and national areas of interest and there is better coverage of publications in SSH disciplines. Second, the databases can provide quantitative evidence to support qualitative peer-review assessments, particularly in research areas not sufficiently indexed in Scopus or WoS. Third, the databases can be co-constructed by the academic and research community and are hence more fair and transparent for research evaluation and other uses. Last, national and institutional databases can also be more open to recommendations of new, open access publications. This study provides some supporting evidence for the vision of a European databases (see Puuska, et al., 2018) by comparing the coverage of the Scopus, WoS, and the OBRSS list.

### Acknowledgements

Dr. Lai Ma would like to thank Professor Gunnar Sivertsen for his generous support during her STSM visit to Nordic Institute for Studies in Innovation, Research and Education (NIFU) in Oslo, Norway. This work has been conducted within the framework of the COST action “European Network for Research Evaluation in the Social Sciences and Humanities” (ENRESSH, CA15137, [enressh.eu](http://enressh.eu)).

### References

- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42, 725–737.
- Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation*, 12(1), 39–46. <http://doi.org/10.3152/147154403781776780>
- Engels, T. C. E., & Guns, R. (2018). The Flemish performance-based research funding system: A unique variant of the Norwegian model. *Journal of Data and Information Science*, 3(4), 44–59. DOI: 10.2478/jdis-2018-0020
- Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, 24(1), 63–77.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261. <http://doi.org/10.1016/J.RESPOL.2011.09.007>
- Ma, L. Responses to Output-Based Research Support Scheme in University College Dublin, 23<sup>rd</sup> International Conference on Science and Technology Indicators (STI 2018), Leiden, The Netherlands, September 12–14, 2018.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106, 213–228. DOI: 10.1007/s11192-015-1765-5
- NSD. (2018). The Nordic List. Retrieved from [https://dbh.nsd.uib.no/publiseringsskanaler/Forside.action?request\\_locale=en](https://dbh.nsd.uib.no/publiseringsskanaler/Forside.action?request_locale=en)
- Ossenblok, T. L. B., Engels, T. C., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science – a comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation*, 21(4), 280–290.
- Pölonen, J. (2018). Applications of, and experiences with, the Norwegian Model in Finland. *Journal of Data and Information Science*, 3(4), 30–43. DOI: 10.2478/jdis-2018-0019



- Puuska, H-M., Engels, T., Guns, R., Sivertsen, G., Pölonen, J., & Jorge, M-R. (2018). Proof of concept of a European database for social sciences and humanities publications: The VIRT-ENRESSH pilot. CRIS2018: 14<sup>th</sup> International Conference on Current Research Information Systems, Umeå, June 13-16, 2018.
- Schneider, J. W., Aagaard, K., Block, C. (2016). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation*, 25(3), 244-256. <http://doi.org/10.1093/reseval/rvv036>
- Sīle, L., et al. (2018). Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. *Research Evaluation*, forthcoming. Retrieved from <https://doi.org/10.1093/reseval/rvy016>
- Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter*, 6(1), 22–28.
- Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science*, 3(4), 2-18. DOI: 10.2478/jdis-2018-0017
- Sivertsen, G. & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics*, 91(2), 567–575.
- van Leeuwen, T.N., van Wijk, E., Wouters, P.F. (2016). Bibliometric analysis of output and impact based on CRIS data: a case study on the registered output of a Dutch university. *Scientometrics*, 106(1), 1–16.
- Zacharewicz, T., Lepori, B., Reale, E., & Jonkers, K. (2018). Performance-based research funding in EU Member States—a comparative assessment. *Science and Public Policy*. <http://doi.org/10.1093/scipol/scy041>