Essays in the Art and Science of Academic Journal Editing and Publishing

Policy Development and the SCOPUS Citation System

Part 4: The Work of the Content Selection Advisory Board 2020 – 2021

David Rew, MA MB MChir (Cambridge) FRCS (London)

Honorary Consultant Surgeon to the Faculty of Medicine, University of Southampton, UK
And to the Clinical Informatics Research Unit.

Former Editor in Chief of the EJSO, The European Journal of Surgical Oncology, 2003-2009

Subject Chair for Medicine to the SCOPUS Content Selection Advisory Board, Elsevier BV, The Netherlands, 2009 to the Present

A Working Paper for publication on the ePrint Server of the University of Southampton

2nd March 2026

Correspondence to dr1@soton.ac.uk

Key Words: SCOPUS, Content Selection Advisory Board, Bibliometrics; Web of Science, Publishing Ethics, Publication Malpractice; SCOPUS Radar;

Contents

Abstract

Introduction

The Covid 19 Pandemic and the Pandemic of Academic Publications

The SCOPUS CSAB Virtual Meetings, May 28th – June 18th, 2020

The Expansion of Access to SCOPUS Content in New Commercial Models

A Review of the SCOPUS Title Re-evaluation Programme

Policies on Preprint Accrual to SCOPUS.COM

SCOPUS support for regional research evaluation: The Japan Cabinet Office Exemplar

Policies on Scopus Core Content Coverage

Policy on the Coverage of Arts, Humanities and Social Sciences Journals

Policies on Optimising Quality Assessment

The Recognition and Management of "Predatory" Journals and Citation Malpractice

SCOPUS Radar and the Detection of Outlier Journals

The Policy on Remote Zoom Meetings

The SCOPUS CSAB Virtual Meetings October 16th to November 20th 2020

The Researcher-Centric Perspective on SCOPUS as a Resource

Trust and Content policies for SCOPUS

Policy on the Timing of the Cessation of Content Feeds for Discontinued Journals

SCOPUS Content Expansion Policies

The Researcher Entity

Policy Work of the CSAB in 2021

The Moscow International Scientific and Practical Conference, 24th-25th May 2021

The SCOPUS CSAB Virtual Meetings, May 18th – July 2nd, 2021

The Policy of Public Engagement about SCOPUS Strategy and Content

The IAP Project- Combating Predatory Academic Journals and Conferences

Policy on the Management of Discontinued Journals in SCOPUS

Policy on References in Roman and Non-Roman Script in Scopus.

The Infodemic of distrust in academic publishing
A Policy on Theses, Dissertations and Proquest
A Policy on Poly-Authorship
A Practical Example of Poly-Authorship

The SCOPUS CSAB Virtual Meetings of October 19th – November 15th, 2021
Policies of Trust and the new SCOPUS Research Data Platform
The Probabilistic Nature of Entity Data in the SCOPUS System
An Operating Model for the Policies and Integrities of Trust
Practical Examples of the Importance of Data Integrities
The Policy on Self-citations
Policy on Modernisation of the Scopus Title Evaluation Platform
The Impact of Machine Translation on the English Language Policy
Our Discussions on Russian <-> English Machine Translation
Our Paper on Machine Translation in Russian Science Editor
The Settled Policy for English Language Use in SCOPUS in 2021
Citation Metrics, Altmetrics and the Actual Usage of Published Papers

In Summary
Acknowledgements

Postscript: Depatures from the Board during the Covid Pandemic
The Departure of Professor Evan Bieske (Chemistry)
The Death of Professor Peter Stambrook

References

Abstract

SCOPUS is a key component of an academic information system which has been developed by the Dutch company Elsevier BV for the provision of information on the performance of authors, journals, books, publishers, faculties and institutions to Universities, Corporations and Governments around the world. It functions in respectful competition with the Web of Science, which is owned by Clarivate Analytics.

Trust and Quality Assurance are central to the societal value of bibliometric systems. The SCOPUS Content Selection Advisory Board (CSAB) was appointed in 2009 to oversee the quality of SCOPUS content through the SCOPUS Title Evaluation Programme (STEP) and to provide independent advice to Elsevier on related matters.

The development of consistent policies has been central to the work of this Board, thorugh a period from 2009 onwards when the entire academic publishing landscape has evolved and changed radically in consequence of :

- The expansion of the Internet and its enabling technologies:
- The development of the open access movement, and the adaptations of publishers to it.
- The transition from paper to web-based technology, and the centrality of the journal website;
- the transition from subscription based funding to “Author Pays” and “Article Processing Charge” (APC) based systems.
- the industrialisation of publication fraud to secure personal and commercial advantage by gaming the methodologies of bibliometrics; and
- the explosive growth of generative Artificial Intelligence technologies .

In previous essays in this series, I have described the work of the SCOPUS CSAB from 2009 to 2019 in the development of policy and related matters. In this essay, I describe the policy and development work of the Board through the Covid pandemic of 2020-2021.

General Introduction

The Covid 19 Pandemic and the Pandemic of Academic Publications

Soon after we flew out of Beijing at the conclusion of the SCOPUS Content Selection Advisory Board (CSAB) meeting in November 2019, the Coronavirus variant **SARS-CoV-2** flew out of Wuhan and around the world in a few weeks. There followed a rolling global societal lockdown. This was the only realistic way to control the spread of this virulent airborne respiratory pathogen in the initial absence of an effective vaccine.

It rapidly became apparent that it would not be possible to hold a meeting in person of the Board in the Spring of 2020. Fortunately, the internet enabled networking systems TEAMS, ZOOM and other tools were rapidly improved to maintain business continuity.

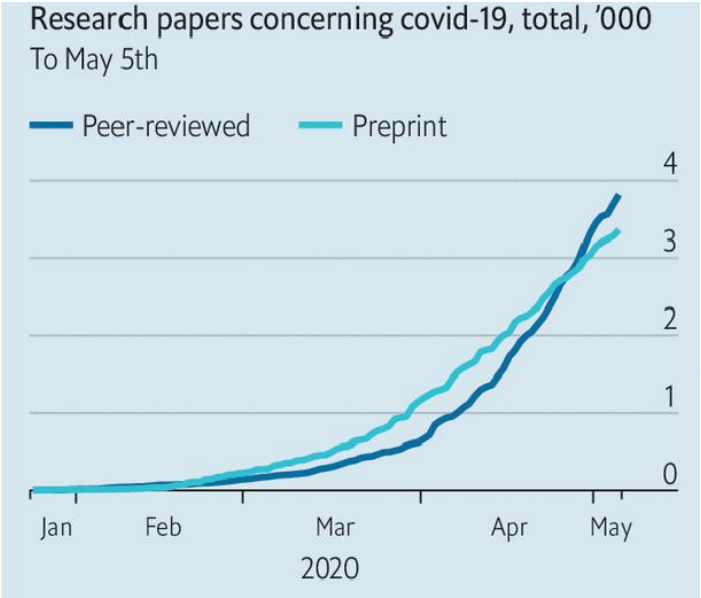


Figure 1 From The Economist May 7th 2020 (see text)

Covid 19 also had a huge impact on submission volumes in academic publishing (Figure 1). The explosion in **Preprint Papers** on Covid-19 was reported in the Science & Technology section of the May 7th 2020 edition of The Economist. Its writers noted that: "Scientific research on the coronavirus is being released in a torrent". More than 7,000 papers on the pandemic had appeared in the previous three months, of which 20% had been uploaded in the previous week.

As *The Economist* went to press in early May 2020 the BioRXiv and MedRxiv servers featured between them, 2,853 articles about SARS-CoV-2 or covid-19. Another 789 had been posted on ArXiv, which specialises in maths and physical sciences. The Economist writers opined that scholarly communication seemed to be at an inflection point.

The SCOPUS CSAB Virtual Meeting via Zoom May 28 – June 18, 2020

Given the global ban on international travel, we organised our first meeting of 2020 as a series of short ZOOM sessions. Over four two hourly sessions between Tuesday, 2nd June and Thursday, 18th June, we discussed a range of topics and their relevant policy implications, including:

- Scopus strategy and the impact of COVID;
- The early indication of research through preprints;
- Scopus capabilities as a service to address regional research;
- Extended datasets, and the needs of bodies such as the Japan Cabinet Office;
- Development of the new CiteScore metric;
- Pro-active title selection & embargo periods;
- Future of title re-evaluation.

On June 2nd 2020, we were introduced to the Elsevier & Scopus strategies to help mitigate the impact of COVID-19. The dramatic increase in COVID-19 manuscript submissions was noted, particularly from China. Some 3000 papers and 1000 review articles had already been received by Elsevier. This volume was likely to overwhelm editors, reviewers and production teams. Computer algorithms were needed to fast track papers in open access journals. Data was also shared with the World Health Organisation (WHO).

Jörg Sack (Subject Chair for Computer Sciences) was concerned that fraudulent published research might spread like wildfire with population-wide impacts; and he highlighted the importance of our work for the Board in mitigating such impacts.

Maxim Khan for **Elsevier** explained the various user cases for the SCOPUS.com data system. He explained that:

- Scopus.com had a researcher focus.
- The Scopus Data and Scopus API products had an institutional focus.
- National bodies were using Scopus data to allocate resources to research teams;
- Universities were making decisions about collaborations and hiring;
- Researchers used it to maintain their profiles and to seek collaborators;
- Publishers were using Scopus to decide which journals to support.

These user cases described the four primary market segments of:

- National level evaluations;
- University benchmarking;
- Individual evaluations and appraisal systems;
- Journal evaluations.

Scopus.com therefore offered a portfolio of functions that help people in academic and corporate research to make decisions. This observation led to a discussion on experimentation with the SCOPUS Application Programming Interface (API) suite of software tools. These would permit the integration of SCOPUS data with private or corporate data sets to produce novel outputs.

The Expansion of Access to SCOPUS Content in New Commercial Models

As a working clinician, I had long felt that SCOPUS contained rich content which might be made available in a commercially practical form to the general non-academic user on a personal or individual basis. I wrote following Max's presentation that:

"The vast majority of Higher Education graduates worldwide do not spend their lifetime in University careers.... However, many will retain a lifelong interest in the subjects of their professional and academic grounding and they are likely to return to search engines on a regular basis for specific purposes.

I suggested that:

The imaginative use and marketing of SCOPUS content on a subject by subject and topic by topic basis, and the use of novel information display tools, might motivate a much wider audience to use SCOPUS content. This in turn would bring the greater rewards to Elsevier from market penetration and new market development to justify further investment.

I therefore feel obliged to continue to raise the case for those hungry millions of "SCOPUS-dispossessed" professionals in non-academic careers who presumably value quality assured information, but who have little or no knowledge or experience of SCOPUS-assured resources in their present employment. I asked whether Elsevier could expand the user cases and spin out models for further for further discussion."

In ensuing discussions, we noted that the open provision of Coronavirus content might identify a clear need and might be a first step to a target model for individuals

A Review of the SCOPUS Title Re-evaluation Programme

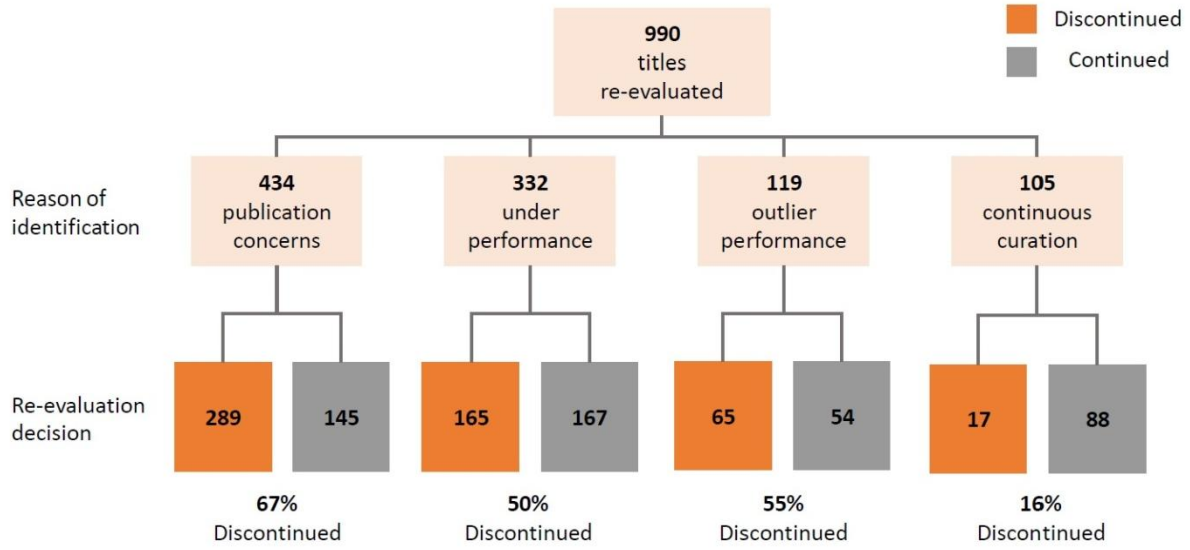


Figure 2. The SCOPUS Re-evaluation Programme statistics 2016-2020 (Elsevier data)

We also discussed progress with the Title Re-evaluation project, through which 990 titles in SCOPUS were investigated for a wide range of reasons between 2016 and 2020, which were broadly classified, as shown in Figure 2.

Policies on Preprint Accrual to SCOPUS.COM

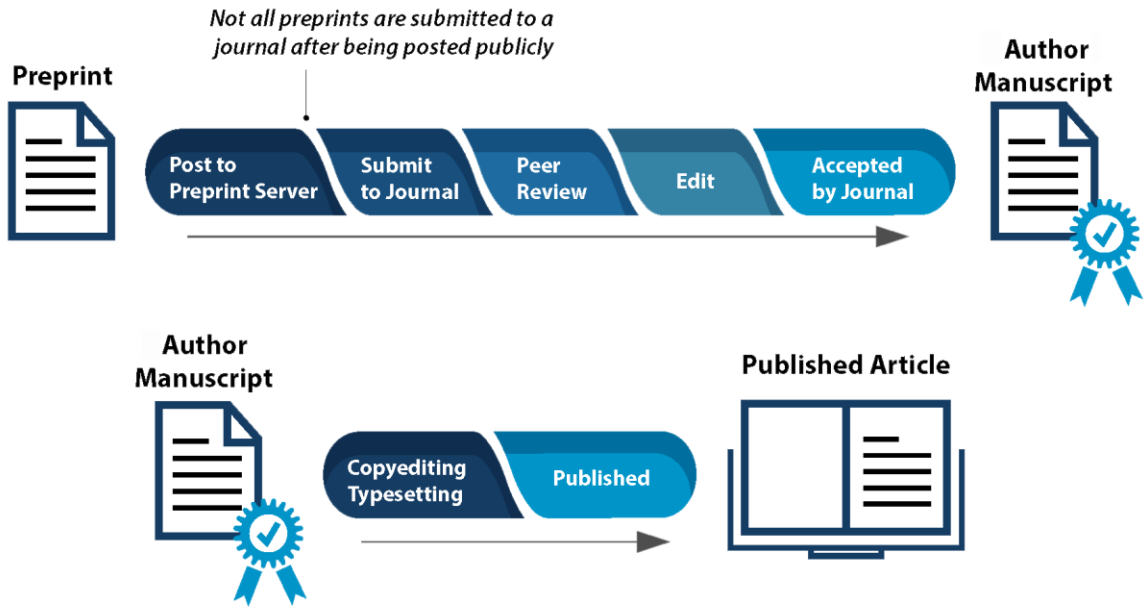


Figure 3. A schematic for the maturation of a Preprint manuscript (Image: NIH website)

Preprints are versions of academic papers which precede peer review. The history of preprints dates back to the 1960s, when the US National Institute for Health (NIH) first accepted their legitimacy in the biological subject fields. Their use expanded dramatically in the early 1990s with the growth of the Internet, and particularly with the development of the TeX file format which simplified internet transmission of documents.

The leading Preprint server ArXiv was founded in the US academic community in ~1990 around the Los Alamos National Laboratory with physics files. ArXiv was formalised as ArXiv.org at Cornell University in 2001, and its coverage expanded to include astronomy, mathematics, computer science, quantitative biology and statistics.

Preprint servers met the market demand for a fast, low cost means to put new material in front of the target readership at the earliest opportunity without the delays or unpredictability of peer review. In many cases and in key subject areas, this facility fulfilled the needs of the authors for rapid and effective public communication, whether or not they intended to progress to formal commercial publication.

In other cases, the expectation was that these preprints would proceed to formal publication by a commercial publisher to "authenticate" the paper (Figure 3.) This would lead to two versions of a paper in circulation: the preprint version on ArXiv or another preprint server, and the peer reviewed version in a formal journal.

This situation was further complicated because **Digital Object Identifiers (DOIs)** were allocated to both forms of publication, this making all versions of documents both searchable and potentially confusing from amendments in the formally published version.

For this reason, preprints were not recognised as being citable in SCOPUS and other systems. However, the rapid expansion of Preprint servers and the increasingly common practice of leaving the preprint server version as the definitive version made this policy of exclusion increasingly untenable.

In 2016, Elsevier had acquired the Social Sciences Research Network (SSRN) preprint server. SSRN dated back to 1994. It was designed to serve as an immediate repository for uploaded and unreviewed papers. It evolved to serve a wide range of disciplines, including the social sciences, engineering sciences, humanities, life sciences, applied sciences, health sciences, and physical sciences. Other publishers followed a similar strategy

The use of pre-print servers expanded to Medicine. The Covid pandemic saw an early flood of papers from Medicine fields to preprint servers, thus further expanding the general awareness of the utility of preprint servers for publication "at the speed of thought". Moreover, the use of "Moderators" to oversee the submissions to preprint servers added a modest layer of defence of quality assurance to the preprint inflows.

We therefore reviewed our earlier Preprint policies for SCOPUS. We recognised preprint servers as formal publishing systems. We recommended the decision to include preprint data in author profiles in SCOPUS, so as:

- To create a clear preprint user experience;
- To enrich researcher profiles;
- To provide earlier indications of research activity;

- To focus on the most recent preprint content (2017 - present); and
- To focus on the major servers, ArXiv, BioRxiv, ChemRxiv, SSRN and MedRxiv.

We supported the Elsevier argument that we should not at this point:

- Mix preprints and published articles;
- Make preprints searchable on Scopus;
- Create new author profiles based on preprint content only;
- Integrate pre prints into Scopus metrics.

It remained to be seen whether this holding position in respect of the principal of peer review and the prevailing publishing business model would be defensible in the longer term.

It was initially estimated that 500,000 preprints would accrue to SCOPUS, and these would enrich 1.1-1.2 million author profiles. Physics and Mathematics would particularly benefit from this decision. Social science preprints from SSRN would include ~250,000 items and ~1.5 million other author profiles would be enriched.

Jörg Sack emphasised the need for disambiguation and matching of different versions of preprints and definitive papers, and that we needed statistics on citations to preprints but not from preprints. I noted the rapid expansion of MedRxiv through Covid submissions.

Michael Lehman (Neurosciences) noted that this topic was particularly important for a grant review, in the context of the focus on Rigour and Reproducibility of the US National Institute for Health. The link of the Preprint to the final version of record is a very important point for funders and for grant review panels. It is important to present assessors with such data in an easy and readily understandable way, to assist funding decisions.

To further confuse the preprint landscape, more than universities have adopted “eprint servers” since the early 2000s, through which locally generated content is published directly to the internet without cost or the delays of peer review.

SCOPUS support for regional research evaluation: The Japan Cabinet Office Exemplar

Mohan Paturi, Professor of Computer Science at University of California, San Diego, described his interests in algorithms, computational complexity and machine learning. Mohan had founded Parity Computing, which was acquired by Elsevier in 2019 to develop data unification and analytics tools, predictive models and customized systems.

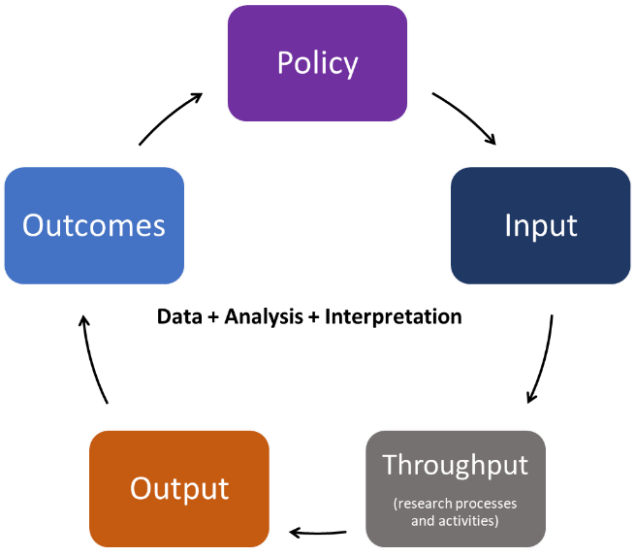


Figure 4: The Feedback Loop between Policy, Data and Analytics (courtesy of Mohan Paturi).

Mohan noted that national research governance bodies have limited resources, so they must be strategic and prioritise resources and design effective policies. He explained his Unification model of a Feedback Loop (Figure 4) between Data, Analysis and interpretation.

He posed the problem that if a country wants to develop a capability in quantum computing, and to understand the basic science and its application to the manufacture of quantum computers, then what are the best strategies in the national context?

Scopus had already collected huge quantities of data, and normalised records through deduplication, linkage, profiling entities and classification of content. This data now needed curation using analytics tools to help answer practical questions through predictive models and to guide future initiatives.

In this context, in March 2020, the Japanese Cabinet Office (JCO) owned some 22 million records from more than 3,600 indigenous publications. Most of which were written in Japanese, and only 40% were listed in Scopus. The Government of Japan sought to promote evidence-based policy making in funding research, in order to create an **Evidence Based Policy Making System (EBPMS)** through the JCO Council on Science, Technology and Innovation. A programme was proposed to integrate Japanese data with global SCOPUS data to better contextualise and understand Japanese national performance.

Japan already had a number of bibliographic databases for Japanese Academic Society publications. CiNii, **the Citation system for the National Institute of Informatics of Japan** was the front-end for the federation of the databases.

The workflow of the proposed project is illustrated in Figure 5. The first step was to build an accurate and comprehensive database of the output of Japanese researchers, based on a proposal to enrich their author profiles on a custom basis for national use.

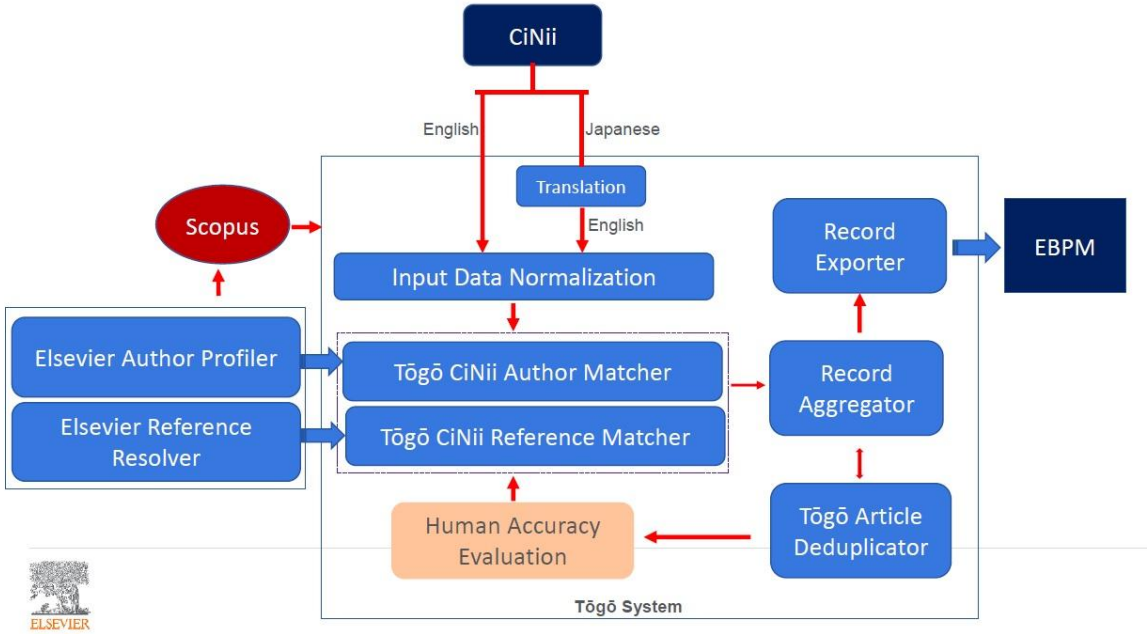


Figure 5: Proposed workflow of a data integration project between SCOPUS and CiNii content (Courtesy of Mohan Paturi)

Mohan provided the live example of an Associate Professor at JF Oberlin University who had some 70 publications recorded in CiNii, most of which were in Japanese script, but only six

of which were recorded in his SCOPUS Author profile. Combining of these two data sets provided a much enriched Author profile (Figure 6).

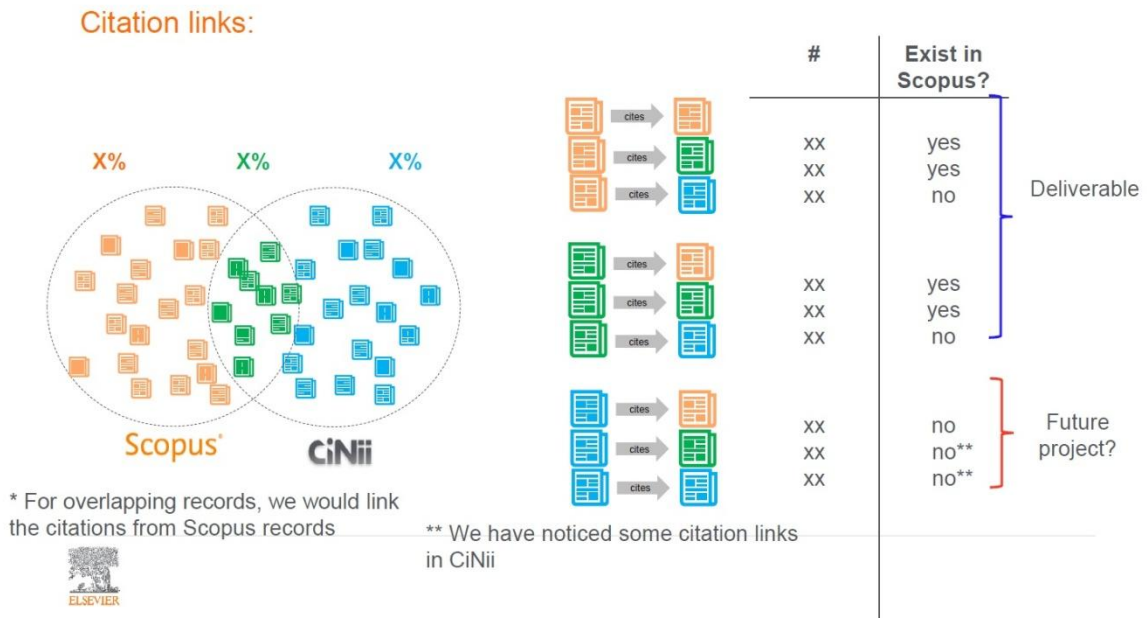


Figure 6. Citation links between SCOPUS and Japanese national data (from Mohan Paturi)

The complexity of merging bibliometric data sets is highlighted by Mohan in this further diagram. For those local citations in CiNii which were not already in SCOPUS (in blue), a further project would be needed both to transliterate references from Japanese to Latin script and to incorporate them in the database.

For each Scopus-listed Japanese researcher, it would be necessary to generate a publication profile that included both Scopus and CiNii document records. For each Scopus document record, it would be necessary to resolve and de-duplicate the references in the record to CiNii articles.

As of early 2026, public internet resources indicate that this project has progressed to the exchange of metadata but understandably not to full integration of CiNii and SCOPUS.

The exemplar nevertheless demonstrates a practical solution of wider application to the integration of the global data set which SCOPUS holds with specific data sets held by individual governments and institutions, with original sources which may be held in languages and scripts other than English and Latin.

Policies on Scopus Core Content Coverage

We considered in our Virtual Meeting on June 15th 2020:

- Updates from the Korean, Chinese, Russian and Thai Local Boards.
- Pro-active title selection strategies to improve SCOPUS coverage;
- The Emerging sources Citation Index (ESCI) of Clarivate Analytics.
- 100 titles in other Elsevier databases (Embase, Compendex, Reaxy) for evaluation;
- The use of the SCOPUS RADAR tool to find highly cited but overlooked titles

Policy on the Coverage of Arts, Humanities and Social Sciences Journals

The Art, Humanities and Social Sciences cover a wide range of content within the All Subjects Journal Classification Scheme, as set out in Table 1. The number of journals which are covered in SCOPUS rapidly expanded from 2010 onwards. We agreed that the oversight of the portfolio would remain within the remit of a small specialist team of Subject Chairs, including Richard Whatmore (History), Henry Yeung (Social Sciences) and David Nelken (Law) (Table 1).

Arts and Humanities	**Social Sciences**
Archaeology (arts and humanities)	Anthropology
Arts and Humanities (miscell)	Communication
Classics	Cultural Studies
Conservation	Demography
History	Development
History & Philosophy of Science	Education
Language and Linguistics	Gender Studies
Literature and Literary Theory	Geography, Planning & Development
Museology	Human Factors and Ergonomics
Music	Law
Philosophy	Library & Information Sciences
Religious Studies	Life-span & Life-course Studies
Visual Performing Arts	Linguistics & Language
	Political Science & Internat. Relations
	Public Administration
	Safety Research
	Sociology & Political Science
	Transportation & Urban Studies

Table1. Subject classification for A and H, and Social Sciences Journals, as derived from the Elsevier All Sciences Subject Classification scheme.

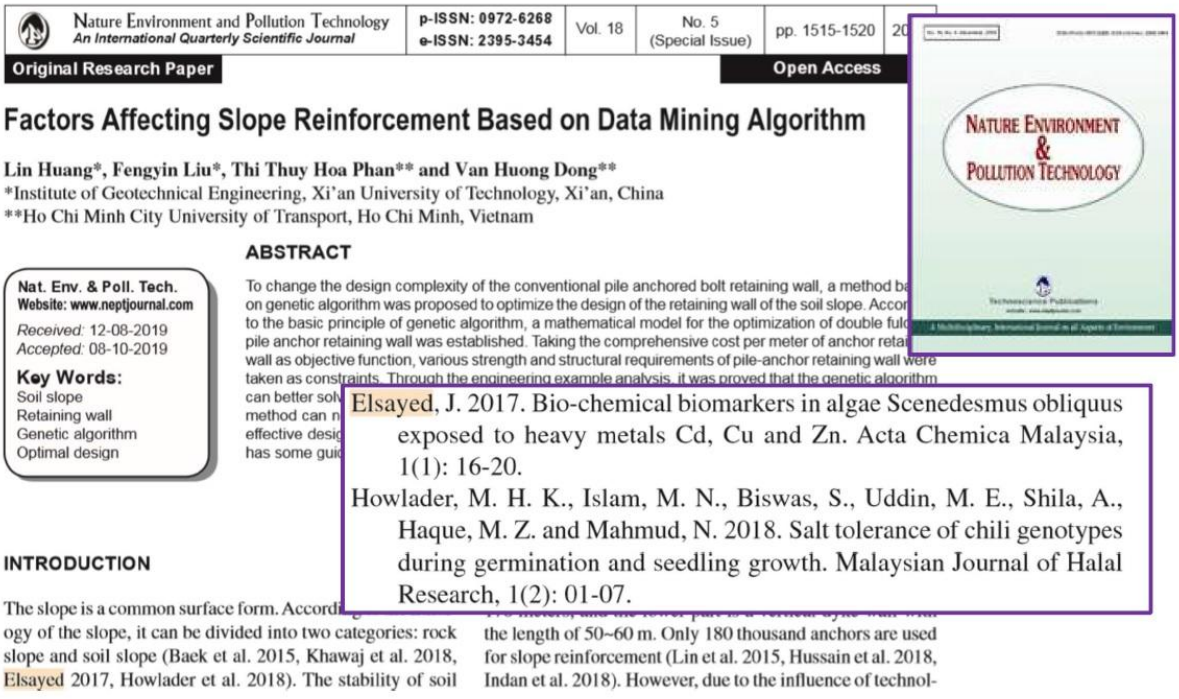
Policies on Optimising Quality Assessment

The Recognition and Management of "Predatory" Journals and Citation Malpractice

During our virtual meeting on June 18th 2020, we again discussed the need for practical and acceptable definitions our focus on predatory publishing, following the professional tribulations of Jeffrey Beall. We were increasingly receiving reports of new forms of publication malpractice, including Citation and Publisher Cartels and citation clusters; the recycling of articles; the citing of unrelated works; untrustworthy publishers and fake editorial boards.

We noted how demanding of time, resources and insights the detection of such fraud would be, and how the prospect of publishing in a SCOPUS listed journal could well be a magnet for such behaviours for unscrupulous researchers, authors and publishers.

Peter Brimblecombe (Environmental Sciences) presented examples of citation malpractice. Figure 7 illustrates a case in which irrelevant references have been inserted into the text. The explanation as to how this fraud may have been committed eluded us at the time, but we now better understand the workings of paper and citation mills, and citations for sale.



Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
e-ISSN: 2395-3454

Vol. 18 | No. 5 (Special Issue) | pp. 1515-1520

Original Research Paper | Open Access

Factors Affecting Slope Reinforcement Based on Data Mining Algorithm

Lin Huang*, Fengyin Liu*, Thi Thuy Hoa Phan** and Van Huong Dong**
*Institute of Geotechnical Engineering, Xi'an University of Technology, Xi'an, China
**Ho Chi Minh City University of Transport, Ho Chi Minh, Vietnam

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com
Received: 12-08-2019
Accepted: 08-10-2019
Key Words:
Soil slope
Retaining wall
Genetic algorithm
Optimal design

ABSTRACT

To change the design complexity of the conventional pile anchored bolt retaining wall, a method ba
on genetic algorithm was proposed to optimize the design of the retaining wall of the soil slope. Accor
to the basic principle of genetic algorithm, a mathematical model for the optimization of double fulc
pile anchor retaining wall was established. Taking the comprehensive cost per meter of anchor reta
wall as objective function, various strength and structural requirements of pile-anchor retaining wall were
taken as constraints. Through the engineering example analysis, it was proved that the genetic algorithm
can better solv
method can n
effective desig
has some guid

Elsayed, J. 2017. Bio-chemical biomarkers in algae Scenedesmus obliquus exposed to heavy metals Cd, Cu and Zn. Acta Chemica Malaysia, 1(1): 16-20.

Howlader, M. H. K., Islam, M. N., Biswas, S., Uddin, M. E., Shila, A., Haque, M. Z. and Mahmud, N. 2018. Salt tolerance of chili genotypes during germination and seedling growth. Malaysian Journal of Halal Research, 1(2): 01-07.

INTRODUCTION

The slope is a common surface form. Accord
ogy of the slope, it can be divided into two categories: rock slope and soil slope (Baek et al. 2015, Khawaj et al. 2018, Elsayed 2017, Howlader et al. 2018). The stability of soil

the length of 50~60 m. Only 180 thousand anchors are used for slope reinforcement (Lin et al. 2015, Hussain et al. 2018, Indan et al. 2018). However, due to the influence of technol-

Figure 7: Citation malpractice: irrelevant citations which have been placed (sold) in the reference list of a paper on an entirely different subject (Courtesy of Peter Brimblecombe)

On 30th May 2020, The Economist newspaper reported on publication malpractice in detail, as illustrated in the following graphics (Figures 8 and 9):

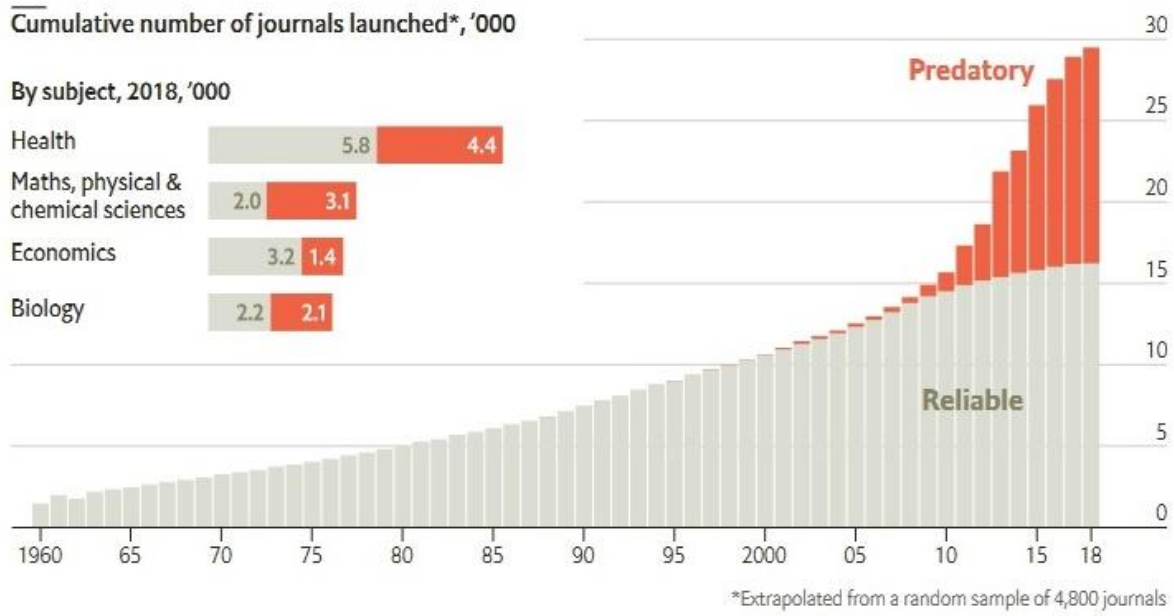


Figure 8: The growth in predatory journals from 1960 to 2018 (The Economist 30.05.2020).

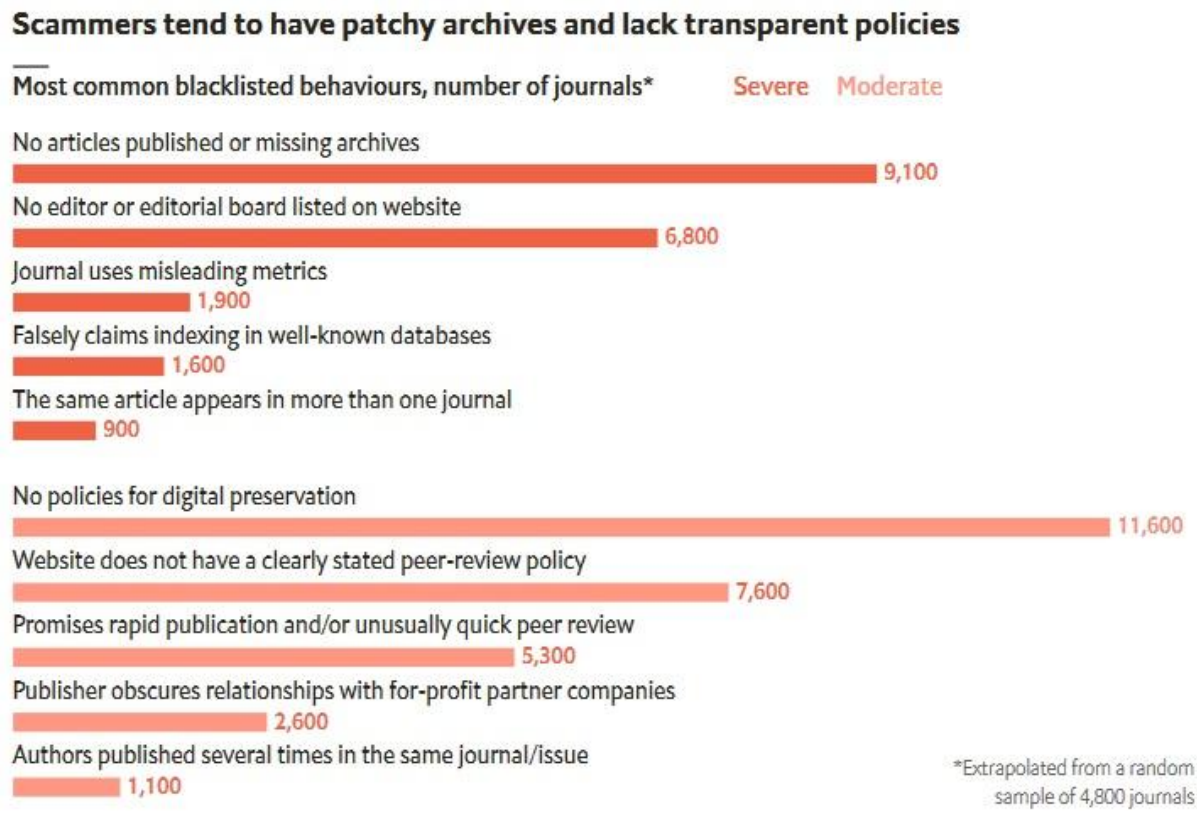


Figure 9: Features of predatory journals (courtesy The Economist 30.05.2020).

See economist.com/graphic-detail/2020/05/30/how-to-spot-dodgy-academic-journals
Figure 9 lists just some of the more easily observed indicators of predatory publication. We now know of many far more complex forms of misdemeanour that frequently evade detection. I note the evidence of the expansion of such journals in the Health Sciences.

SCOPUS Radar and the Detection of Outlier Journals

We were increasingly helped in the detection of malpractice by early iterations of the SCOPUS Radar. These were initially focussed purely on outlier metric measurements, including self citations and total citations. However, we were increasingly receiving reports:

- Of journals which had developed "predatory" characteristics within SCOPUS;
- In the news feeds and academic blogosphere about journal malpractices;
- Of concerns which were reported directly to us by individuals.

By mid 2020, the SCOPUS RADAR had evolved to look at 70 features of all SCOPUS listed journals on a regular scan of the entire database, including features such as the indexing status of suspect journals in other public databases.

We discussed the need for further iterations of SCOPUS RADAR. We needed better tools for the individual subject chairs easily to investigate their own suspicions about journals, authors, editors and publishers, under the generic user case that:
"As a Subject Chair or Independent Researcher, I would like to be able to examine all available information on any title on demand, including its past history of evaluations and its various performance metrics over time, so as to investigate suspicions of malpractice and to determine the case for formal re-evaluation .

In a timely paper in BioRxiv, Andrea Cortegiani and colleagues addressed concerns about the persistent reputational benefit to authors of fraudulent journals, even after their fraud had been recognised and their papers had been censored. They described "**Inflated citations and metrics of journals discontinued from Scopus for publication concerns**" in their " GhoS(t)copus Project" (Andrea Cortegiani et al 2020). They observed that:

"Journals in Scopus are periodically re-evaluated and some are discontinued for publication concerns. However, these journals remain indexed and can be cited. They surveyed the list of 317 discontinued journals for publication concerns from Scopus of July 2019.

294/318 journals published using an Open Access model. The subject areas with the greatest number of discontinued journals were Medicine *(52/317),* Agriculture and Biological Science *(34/317, and* Pharmacology, Toxicology and Pharmaceutics *(31/317). 72/317 were included in the Cabell's blacklist.*

The mean number of citations per year after discontinuation was significantly higher than before (median of difference 64 citations, p<0.0001), and so was the number of citations per document (median of difference 0.4 citations, p<0.0001).

They expressed concern that this paradoxical trend could inflate personal metrics and underwrite career advancement, bonus systems and promotion for authors who were publishing in fraudulent journals. They recommended that countermeasures were urgently needed to ensure the reliability of Scopus metrics both at the journal- and author-level

It was becoming clear that the time was now ripe for a major modernisation of STEP, which had been built in the more trusting academic publishing world of a decade ago, in order:

- To integrate outputs from SCOPUS Radar;
- To provide a much richer picture of individual journal performance;
- To better understand Publishers and their history;
- To better understand author, editor and reviewer behaviour;
- To identify, address and pre-empt sophisticated fraud and malpractice.

The Policy on the Continuation of Remote Zoom Meetings

Through the functionality of the online Zoom meetings, the Board remained coherent and effective through the Covid pandemic. The newly optimised ability to converse worldwide with no perceptible lag in the teleconferencing was remarkable. It was clear that while we were hostage to Covid 19, we had established a precedent for continuing Board meetings on

Zoom in the short to medium term despite the significant limitations and two dimensional remoteness of the format.

Jayarani Raju (Subject Chair for Information Sciences) observed that the format annulled the benefits of the conversational freedom, creative thought and non-verbal signals that emerged from informal one-on-one engagement during face to face meetings. 'Zoom fatigue' and the distractions of multitasking while attending a virtual meeting, also needed to be considered.

The collective view was that the four ZOOM sessions could not satisfactorily substitute for in-person meetings in the longer term. The ZOOM meetings tended to stifle discussion, and to discourage questions. Often, critical voices and ideas were not heard, leading to rapid convergence which often does not lead to the best decisions. The opportunities to talk informally around any one discussion were very limited and the chat facility was unsatisfactory for this purpose.

The SCOPUS CSAB Meetings, October 16th to November 20th 2020

Professor Michael Lehman of Ohio State University succeeded Peter Stambrook on Peter's recommendation as the Subject Chair for Neurosciences. We discussed policy and strategy in respect of:

- Scopus Interface re-design & development roadmap
- **SCOPUS** Content policies
- Data-driven title evaluation and New content types
- Re-evaluation & Publishing Ethics

We discussed the ongoing contributions of the Board, individually and collectively, to Scopus strategy in support of the five key user groups: academic researchers; academic librarians; research leaders; funders; and commercial organisations. We noted the growth of new bibliometric commercial offers, including Semantic Scholar, Microsoft Academic Search, and Google Scholar, which challenged established business models. However, research leaders were still seeking trusted quality assurance signals from SCOPUS.

The Researcher-Centric Perspective on SCOPUS as a Resource

Olivier Dumon of Elsevier explained how the operational emphasis around SCOPUS was evolving from current APIs and user interfaces towards research and knowledge discovery in the data resource. Trusted authority, the continuing expansion of the coverage, the accuracy of the data points, and the ease of use of SCOPUS were all important to this strategy.

We examined the researcher-centric view of the world, and the design of whole-of-life career data profiles which might also capture early career Dissertations, PhDs and Theses, Patents and other outputs across a range of employment groups.

For example, Olivier pointed out that he did not exist as a researcher in Scopus, but that he possessed 11 patents from his work in digital commerce. Jo Angouri (Subject Chair for Linguistics) noted the risk of inadequacies in personal profiles in SCOPUS in A&H and Social Sciences subject fields, where many outputs were not captured by the prevailing SCOPUS collection policies. We therefore considered the research indicators and data sets which were not yet addressed in SCOPUS.

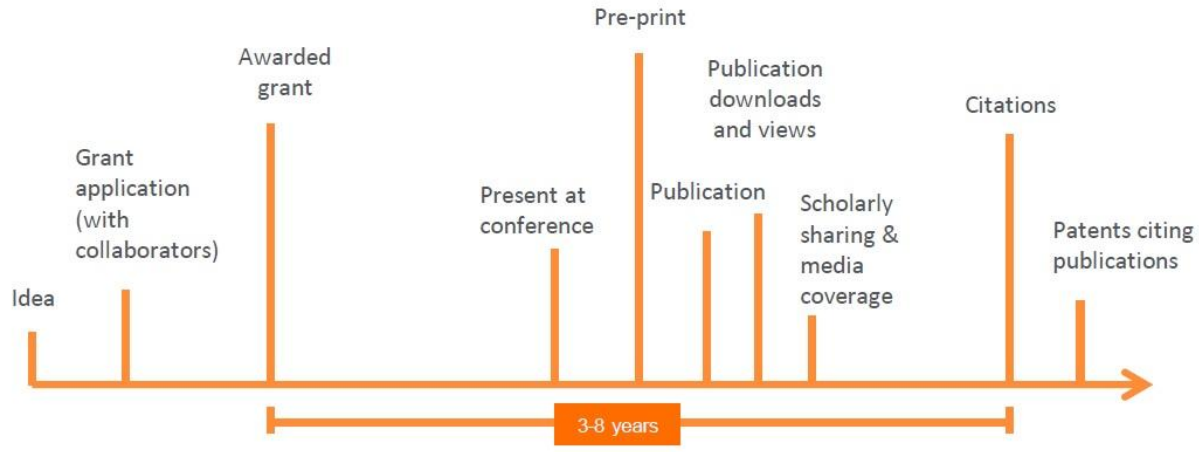


Figure 10a: The early career pathway in respect of impact and metrics (see text)

In this context, Marcel Vonder, Senior Product Director for SCOPUS, mapped out the early career pathway of a typical researcher, following the completion of his or her primary Dissertation, Masters or Doctoral Thesis (Figure 10a).

Marcel then considered how to enrich the researcher profile, using data within SCOPUS, including grant records, references to preprints, SCOPUS view counts and other “altmetrics” (Figure 10b).

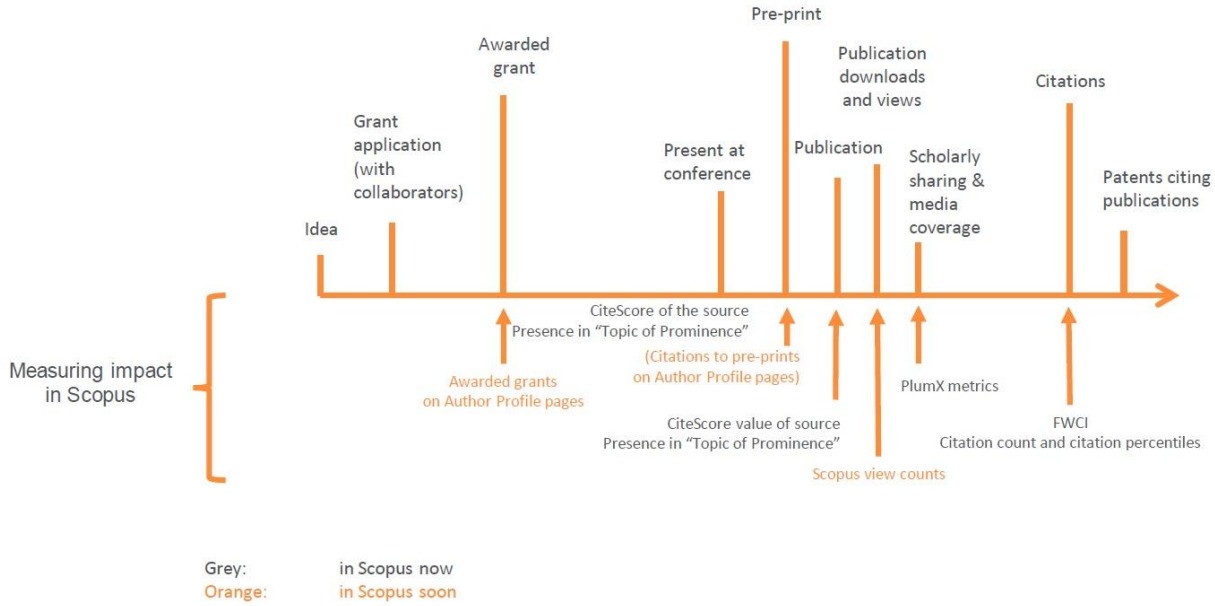


Figure 10b The enriched early career pathway using enriched SCOPUS data (see text)

We also noted design improvements to the SCOPUS interfaces and the development of a mobile-friendly interface, driven by detailed analysis of SCOPUS website usageand market feedback.

Trust and Content policies for SCOPUS

On Monday, 9 November 2020, we considered the importance of Trust in Scopus data, and the contribution of a clear content coverage policy to that Trust. This mandated completeness of accepted content and its linkage with high accuracy to the appropriate entity (author, institution), and its measurement with a relevant indicator.

Content policies address both the selection of desirable content and the exclusion of unwanted content. Wim Meester of Elsevier described how every Document Type should have its own content policies, and how policies start with a clear commercial use case, from which the specific content needs are selected and sourced. Policies had been formally set out for Journal Curation, Medline titles, Discontinued Journals and Books, and they were in

preparation for Conferences and Preprints. The formalisation of SCOPUS Content Policies mandated our review of policies on discontinued journals, and on content expansion.

Policy on the Timing of the Cessation of Content Feeds for Discontinued Journals

Journals were now being discontinued through our re-evaluation programme for a variety of reasons, including for evidence of predatory behaviour and publication malpractice. The discontinuation of the content of a journal, if retrospective, would purge fraudulent articles, but it would also be harsh on legitimate authors who had published in the discontinued journal in good faith. Retrospective purging also caused substantial difficulties in respect of changing existing metrics and a cascade of adverse consequences.

We agreed that the content flow of the journal should be frozen from the moment when a journal is flagged for re-evaluation, along with a public statement and explanation of the reason for discontinuation.

SCOPUS Content Expansion Policies

Wim Meester noted that the citation impact of authored articles is not necessarily a good proxy for professional achievements. He therefore proposed the extension of the scope of the CSAB collection to add more journals and new content formats such as clinical trials, funding sources, as signals as to what the researchers are actually doing (Figure 11). All policies would be made visible to Scopus users, as transparency and good communication of any changes to the policies were very important.

The Researcher Entity

To this point, the information on Scopus.com had been document-driven. Information was extracted from documents in SCOPUS to create the entity called 'researcher'. These profiles would benefit from enhancement with new content types.

On Tuesday, 10th November and Friday, 20th November 2020, Rob Schrauwen of Elsevier expanded on the theme of new content types (Figure 11). He noted that Preprints had now been added to Scopus as a clearly demarcated content type. A plan was in place to include Awarded Grants, Patents, Policy Documents and other measures of Societal Impact.

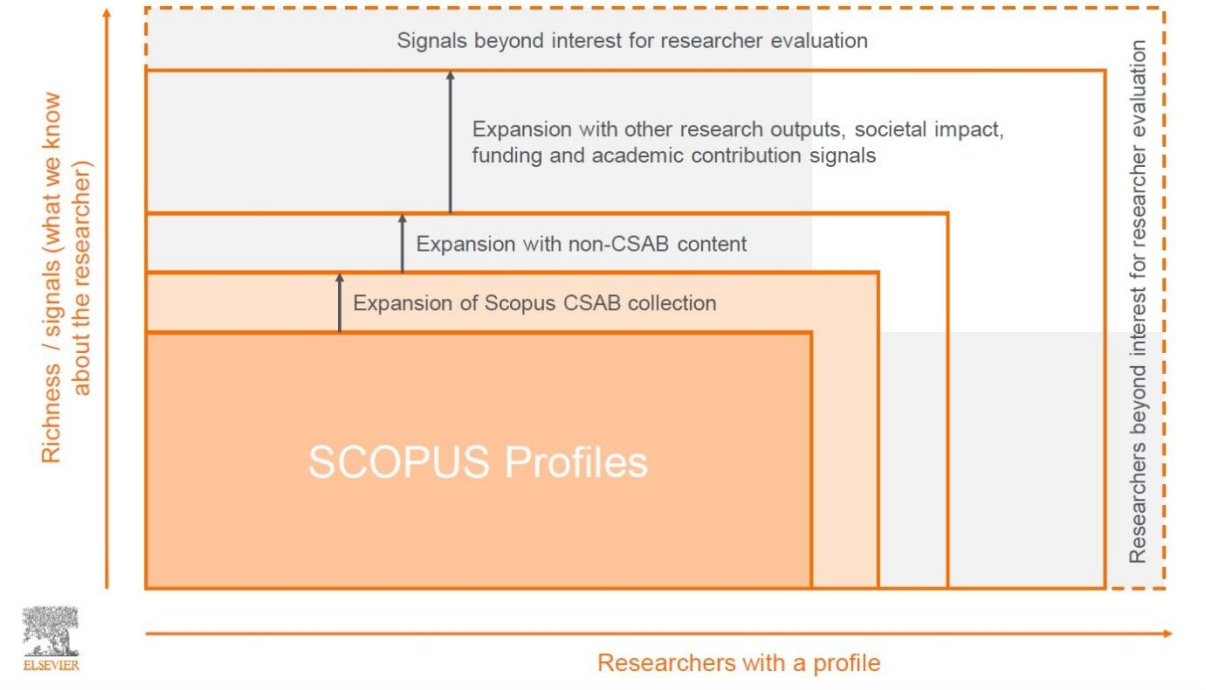


Figure 11. A diagram of the proposal for the expansion of SCOPUS content to create richer researcher profiles with extended data sets (see text) (Image courtesy of Rob Schrauwen)

Rob observed that there were many influential academics whose prominence extended beyond published documents, including:

- Researchers who were publishing in non-SCOPUS listed journals and other sources;
- Researchers in countries with low SCOPUS coverage (eg Japan);
- Editors of journals, books and conference proceedings;
- Academics who publish primarily on preprint servers;
- Those in receipt of prestigious prizes or grants;
- Chairs of learned societies;
- Those who shape high level strategy and policy.

Rob concluded that to improve researcher evaluation, we needed to aim for greater inclusivity of researchers across the academic spectrum, with richer signals and wider coverage within and beyond the CSAB curated collection, and with transparent policies as to what and what not to include in the enriched data set.

In discussion, board members identified a range of policies for further work, including:

- Wider Content Coverage (Books, Book Series; Non-English language journals;
- The role and balance of citation metrics in accept/reject decisions;
- The evaluation of multidisciplinary Journals and Mega-journals for Scopus;
- The evaluation of journals with overt political and propagandist themes;
- Decisions on journals with duplicate or near duplicate titles to other journals;
- The need for greater granularity in the Academic Journal Classification Scheme;
- The identification and management of predatory journals and publishers.
- The need for modernisation of the STEP system;
- Appeals against inclusion/exclusion decisions;
- Reference capture for supplementary materials .

Policy Work of the CSAB in 2021

The Covid pandemic extended into 2021. The arrival of reliable vaccines was beginning to blunt the worst effects and risks of the disease, and vaccination was widespread among health professionals in the UK by the end of January 2021. There was considerable optimism about the restoration of in person meetings during 2021, but it was becoming clear that Covid waves could well continue to resonate globally through the Winter 2021 into the Spring of 2022, and to force continued restrictions which would make travel miserable, where-ever we wished to meet. We were therefore obliged to plan to continue with the virtual CSAB Meetings through 2021.

The Moscow International Scientific and Practical Conference, 24th-25th May 2021

In keeping with our policy of International engagement, Prior to the Spring Board meeting, Tracy Chen of Elsevier and I were invited by Russian Local Board lead Dr Olga Kirillova to speak to a hybrid meeting in Moscow on the theme of "World Class Publication: Global Trends and National Priorities" (Figure 11). The goodwill which flowed from this and other such interactions between the SCOPUS Board and the Russian academic community contrasted with the tragedy of events which were to unfold in and around Ukraine within the year, and the abrupt termination of professional communications which followed.

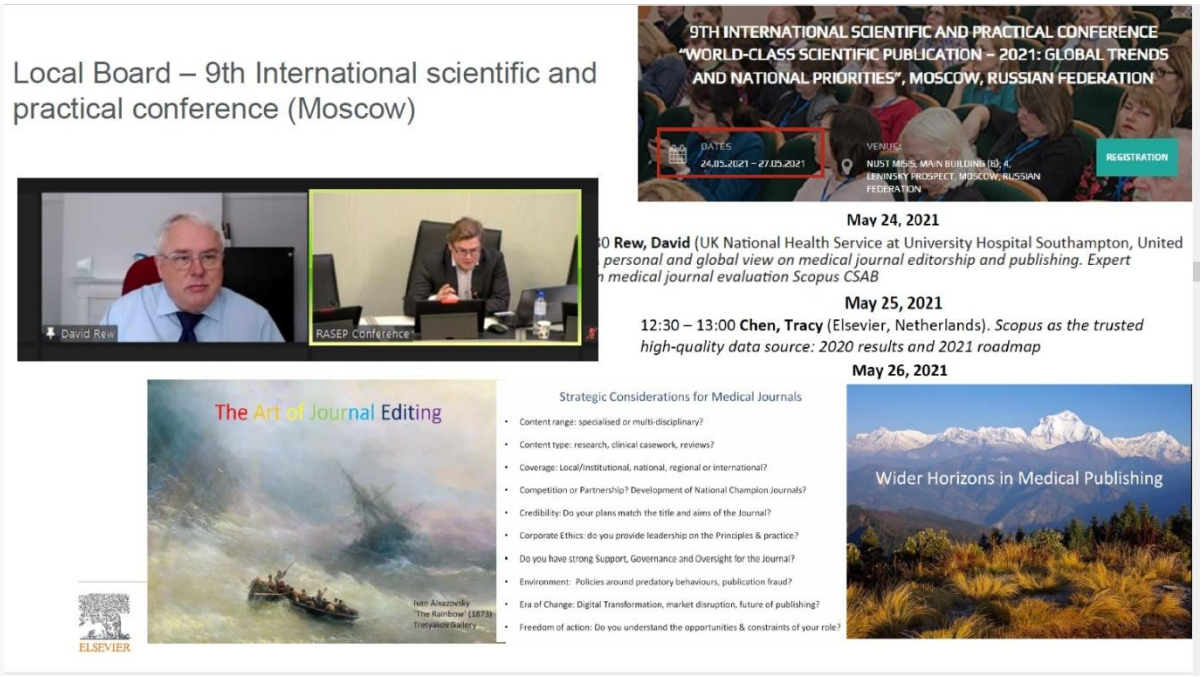


Figure 12. Images from the hybrid Moscow meeting of 24th May 2021

Olga also kindly published a series of my papers in the journal Russian Science Editor and Publisher, as listed in the Reference Section.

The SCOPUS CSAB Meetings via Zoom, May 18th – July 2nd, 2021

Over four sessions, we discussed policies in relation to:

- A strategic vision for Scopus as a vehicle for trusted high quality content;
- The InterAcademy Partnership (IAP) project;
- Combatting Predatory Academic Journals and Conferences;
- Citescore and journal performance;
- Content policies and trends in data, machine learning, and analytics;
- SCOPUS RADAR and the re-evaluation of legacy journals;
- Multilanguage machine translation and its impact;
- Elsevier strategy and corporate responsibilities.

The Policy of Public Engagement about SCOPUS Strategy and Content

The independent global commentariat on bibliometric issues was an important source of insights into problems and challenges within the SCOPUS system. Opinions, informative articles and detailed research into publication fraud were published both in the formal literature and in academic blogs, of which Retraction Watch and The Scholarly Kitchen and noteworthy for their quality.

We discussed in detail a paper by Vit Machacek and Martin Shrolek which had appeared in Scientometrics, which had prompted a report under the headline in Nature News that "Claimed hundreds of predatory journals indexed on Scopus".

The authors had *"mapped the infiltration of journals suspected of predatory practices into the citation database Scopus and examined cross-country differences in the propensity of scholars to publish in such journals".* They had used Beall's lists to obtain the ISSNs of 3,293 journals from Ulrichsweb. 324 journals from Beall's lists with 164K articles over 2015-2017 also appeared in Scopus.

The analysis was demonstrably flawed, as Beall's list had not been maintained since 2017. Beall had worked with a binary classification of predation, whereas "predatoriness" is a complex and nuanced phenomenon. We had already re-evaluated all titles in Beall's list before 2017, and 65% had been discontinued from Scopus. we did not always come to the same conclusion as Beall. The paper was subsequently retracted. (See Macháček 2021):

Wim Meester demonstrated how the paper had nevertheless led to a positive public narrative for the STEP process, in that:

- A full statement was provided to a Nature journalist for publication.
- The sentiment on Social Media became positive as the messaging was reposted;
- Five explanatory webinars had reached ~10,700 attendees.

The IAP Project-Combatting Predatory Academic Journals and Conferences

The InterAcademy Partnership (www.interacademies.org/) represents ~150 institutions and ~ 30,000 members in STEM disciplines. It had become concerned about predatory practices in international publishing. It had defined the spectrum and feature of malpractice from the highest risk to the least risk journals, in the process of which it defined as series of "typical markers" to help define the risk at each point in the spectrum, as illustrated in Figure 13.

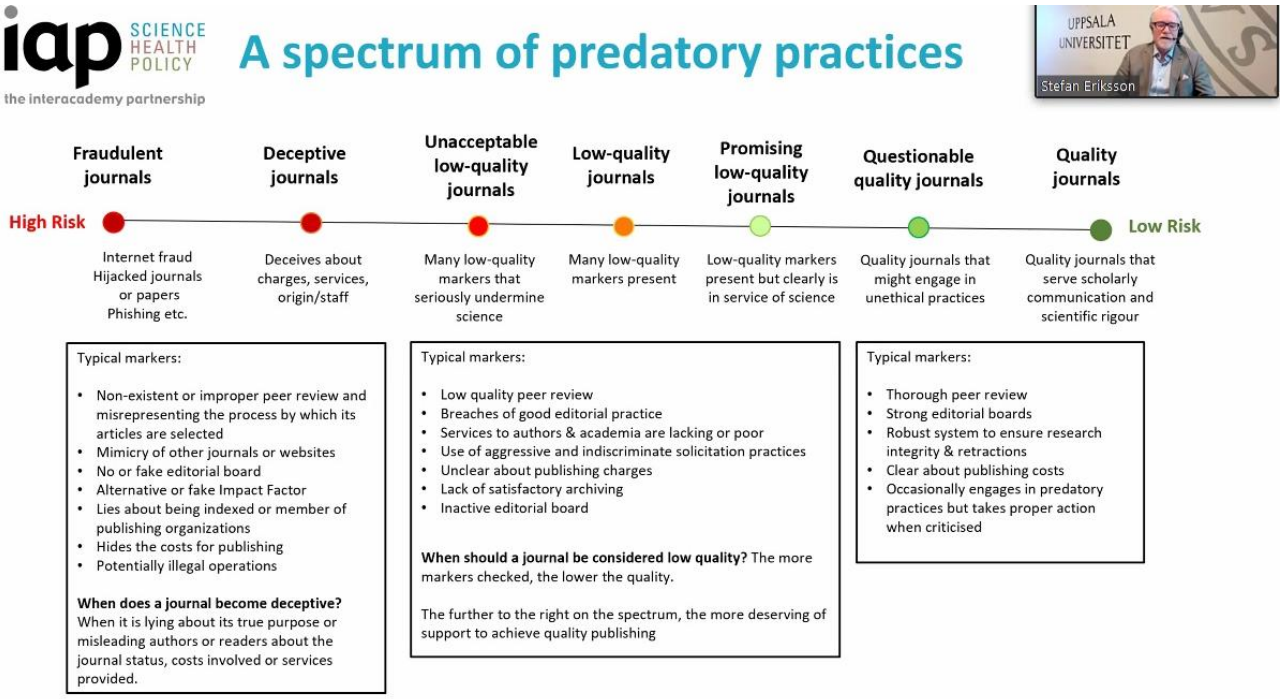


Figure 13. The Spectrum of Predatory Practices (Courtesy of Stefan Eriksson, IAP)

Figure 14: Selected responses to the IAP survey on the use of predatory conferences and journals.

The IAP had conducted an empirical but wide ranging survey of researchers to inform strategies to combat predation, and to raise awareness across research communities. They discovered that at least 25% of respondents had published in a predatory journal or participated in a predatory conference, with comments such as “you need to publish in a predatory journal to stay in the race” (Figure 14).

Figure 15: responses to the IAP survey on the anticipated impacts of predatory publishing

Most respondents believed that predatory practices would have detrimental effects on research if left unchallenged (Figure 15). The preliminary conclusions were as follows.

- Predatory journals and conferences were becoming more sophisticated;
- They were permeating many geographies, disciplines and career stages;
- Post-pandemic working practices might exacerbate predatory practices;
- Distinguishing quality within the spectrum of predation is difficult;
- Raising awareness of the problem is essential;
- Commercial interests of publishers were perceived to be a major challenge;

Somewhat idealistically, the IAP team recommended that publishers should:

- Always put academic excellence before business interests;
- Show transparency and open policies in evaluation and peer review processes;
- Keep Article Processing Charges as low as possible;
- Not create cascade journals to publish material that had been rejected elsewhere.

They observed that many institutions and some nations use SCOPUS as a de facto whitelist, thus putting a great responsibility on SCOPUS and the CSAB to exclude predatory titles .

The Policy on the Management of Discontinued Journals in SCOPUS

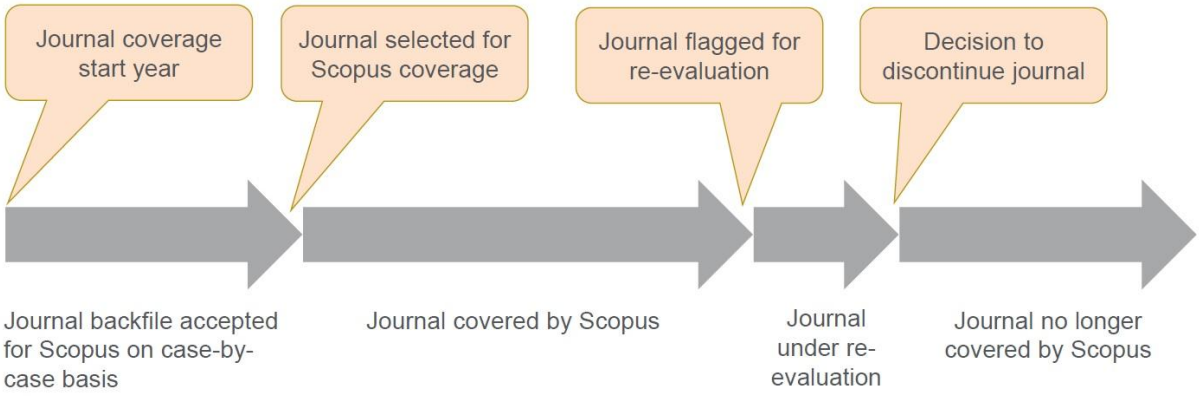


Figure 16: The lifespan of a discontinued journal in SCOPUS

Some journals have a limited lifespan in SCOPUS (Figure 16). In the matter a clear and consistent policy for journal discontinuation, we agreed that Scopus would make no distinction between predatory and otherwise discontinued journals, with no further sub-classification of reasons for exclusion.

For such journals, we explored the options of 'Leaving an archival trace in SCOPUS' *versus* 'Taking everything out of SCOPUS" from discontinued journals. We agreed that:

- The decision to place titles 'on hold' during investigation was supportable;
- For discontinued journals, the 'on hold date' would be the discontinuation date;
- For continued journals, coverage would continue from the 'on hold date';
- A marker was needed in SCOPUS to highlight discontinued journals.

For discontinued journals, content of the journal which is already indexed in Scopus until the moment of re-evaluation would not be deleted and would remain in Scopus, as technical removal would be very disruptive.

This would nevertheless leave the risk that content that had contributed to the de-listing of a journal would nevertheless remain active within the SCOPUS database. We were also advised by the technical team that there would be no filter for users to exclude all content from discontinued journals during Scopus searches.

Policy on References in Roman and Non-Roman Script in Scopus.

Until 2018, references that were not in Roman script had not been captured in the Scopus database, because of the difficulties in handling other scripts, such as Arabic, Chinese, Thai Cyrillic or Korean scripts. The requirement for references in Roman script had been an entry criterion for Scopus review, whereby the STEP team checks a selection of documents for at least ~80% of the references in Roman script.

Up to 75 journals per annum were rejected for SCOPUS between 2010 and 2020 on this basis, amounting to more than 250 journals over that period. With evolving technology, The citation linking for those references had been limited through infrequent references, and because accurate algorithms were not in place to support high quality linking. This limited the ability to develop rich author profiles for researchers outside the Anglosphere.

Wim Meester therefore discussed the proposal to remove the constraint of the criterion for restricting references to Roman script for journals that are considered for review.

A decision was needed because a revised policy would open up SCOPUS to consideration of journals that were previously excluded from SCOPUS on linguistic and technical grounds. Up to this point,

- Titles for CSAB review had translated or transliterated references;
- Scopus captured non-Roman script from the title, name, affiliations and references;
- Not all Non-Roman script titles, references and authors were searchable on Scopus;
- Non-Roman script titles were not counted in profile matching algorithms;
- Local language journals were struggling with translation and transliteration.

Scopus could now capture, process and display major non-Roman scripts, including Chinese and Russian characters. However, the SCOPUS computer systems needed a sizeable corpus of articles with names in both English and local script to permit more precise name disambiguation. The recommendation was endorsed to remove the requirement for references only in Roman script as analytical technology was better able to interpret non-Roman references.

We agreed to continue with a test and development programme to underwrite a new policy on non-Roman scripts. This discussion also highlighted the wider issue of the policy of English language primacy in SCOPUS, given the rapid development of real time and reliable translation systems on websites and for the full text of articles.

The Infodemic of distrust in academic publishing

Elsevier's new Chief Executive, Ms Kumsal Bayasit, concluded the Spring 2021 Meeting with an overview of her ambitions for the SCOPUS Research Data Platform. She described an Infodemic of distrust in academic publishing. An Infodemic is defined by the World Health Organisation as an excessive amount of information, both true and false, about a disease outbreak. An infodemic spreads rapidly and can lead to confusion, risk-taking behaviours, mistrust in health authorities, and an ineffective public health response.

This academic infodemic translated into evidence of distrust in all information sources and otherwise reliable spokespeople, as measured on the Edelman Trust Barometer (see https://www.edelman.com/trust/2021-trust-barometer).

Edelman in a global advisory conglomerate which had cited the Covid19 pandemic as an exemplar of a concurrent Infodemic. In its 2021 report, it noted a failing ecosystem of Trust that was unable to confront the rampant infodemic, leaving businesses, governments, NGOs and the media in an environment of information bankruptcy and with a mandate to rebuild trust and to chart a new path forward.

In this context, Kumsal set out a series of practical objectives within our scope of influence to support Elsevier in addressing the Infodemic, as set out in Figure 18.

Maintain high quality entity profiles (researchers and institutions)	Researcher profile ***accuracy*** improvements
	Institution profiles ***accuracy*** improvements
	Support user feedback for new content types ***(preprints, grants)***
Make our data more versatile	Enable seamless organization lookup (microservice)
	Build a subject classification(s) engine (SDGs, MOE)
Bring new insights for a more complete view of the researcher	Extend the preprint integration into Researcher profiles
	Integrate awarded grants into Researcher profiles
Create new funding insights	Improve coverage and quality of funding opportunities
	Create awarded grant to publication links
	Create awarded ***grant to topic*** links
	Create the best view of ***funding data*** of an article
	Establish quality of ***awardee institution profiles***
Generate insights about the application and impact of research	Assess new content types + their relationship to research output
	Datasets for ***SciVal Societal Impact Module*** (policy, twitter, media)
	Telescope (Research Infrastructure) data – collab w RCO DS
Best in class regional data & open science	Integrate ***Japanese J-STAGE journals*** with Scopus for Japan CaO
	Research data sets institution profiling

Figure 18: Objectives within SCOPUS to create high quality data in the context of the academic publishing "infodemic".

A Policy on Theses, Dissertations and Proquest

In Jul 2021, I digitised my 1991 Thesis, and uploaded it in full to the University of Southampton ePrint server with a doi number, through the help of Paula Head, my faculty librarian. This had led me to review the Proquest website, which partnered with the US Library of Congress, which claims around 9M Theses and Dissertations; and to the UK British Library Ethos Server, which claims around 500K listings. To my surprise and pleasure, as of early 2026, this 1991 Thesis has secured >160 full downloads, thus reinvigorating work that would otherwise have been all but forgotten but for digitisation and the Internet.

We had not seriously considered integration of Theses and Dissertations with SCOPUS, even though they are a rich source of cross-citations and the source of most academic careers. I therefore wondered whether there was an opportunity for an enrichment partnership with organisations such as Proquest.

Ashok Raina (Maths) noted that Proquest originated in University Microfilms at Michigan (Ann Arbor) in the 1950's and 1960's, as a source of copies of PhD theses from American Universities. He thought that this could be replicated in a modern digital form in SCOPUS. He also pointed to the French free service TEL (Theses-en-Ligne) tel.archives-ouvertes.fr/, which was linked to the French version of ArXiv: https://hal.archives-ouvertes.fr/

Michael Lehman observed that Neuroscience dissertations from Dutch PhD programs often took the form of books and they were of very high quality. However, there were significant differences between countries in the quality, format and usefulness of dissertations.

I also recalled Olivier Dumon's early vision for a "Whole of Career" information integration system for Elsevier using Mendeley. In this context, the Thesis becomes the fulcrum for mapping an individual's subsequent career. I hoped that this line of thinking would continue to evolve into a practical solution within SCOPUS.

Regrettably, shortly afterwards, we learned that Clarivate Analytics had beaten us to it with this line of thinking and had acquired PROQUEST for the Web of Science in late 2021.

A Policy on Poly-Authorship

The number of authors per paper has been increasing for many years, both for honest practical reasons and increasingly frequently for fraudulent gain. In some cases, it can amount to dozens or even hundreds of contributors.

Stuart MacDonald of the University of Leicester has asked in the LSE blog “whether authorship mean anything when academic papers are simply citable tokens”. In his 2023 paper in the journal Social Science Information, he states that:
“Many authors in medicine have made no meaningful contribution to the article that bears their names, and those who have contributed most are often not named as authors. Author slots are openly bought and sold. The problem is magnified by the academic publishing industry and by academic institutions, pleased to pretend that peer review is safeguarding scholarship. In complete contrast, the editors of medicine’s leading journals are scathing about just how ineffectual is peer review”

A Practical Example of Poly-Authorship

In early August 2021, John Beigel of the National Institute of Allergy and Infectious Diseases (NIAID) of the US National Institute for Health (NIH) wrote directly to me to raise the issue of how his collaborators should be given credit. John noted that many of his group's papers have a set of authors and a group name. He stated that the group usually has many more authors. In PubMed, members of this group are listed as collaborators to ensure that all who contributed are recognised in PubMed. See pubmed.ncbi.nlm.nih.gov/32445440/

John understood that Scopus records names on the masthead reference, but not the named collaborators, as in PubMed. Numerous people contributed to his Remdesivir study, but the collaborators were not recognised in SCOPUS. He asked whether Scopus would consider listing all collaborators, as in PubMed, so that all members of the "study group" received credit in the h-index calculations.

I passed the question on to Wim Meester for Elsevier, noting that:
- it was clearly in the interests of the authors and of SCOPUS to have maximum author capture, when each of the named collaborators has effectively been validated, and:
- It made sense to be fully aligned with PubMed in these matters.

I therefore wondered whether this placed an onus on the lead authors of any one paper to provide as much information as possible about each co-author and collaborator to simplify data capture, as for example, by ensuring that every contributor had an ORCID identifier.

This also brought us to the principle which was espoused by the Committee on Publication ethics (COPE) that every named author should have demonstrably contributed to the writing of the paper. The discussion highlighted the grey zone between Authorship and Acknowledged Contributors. However, in the real world, individual triallists and groups were contributing to fast moving and large scale trials in the reasonable expectation of academic recognition, so pragmatism was needed in the matter of attribution.

It also raised the question of how many author listings is enough, given that some of the big projects in high energy physics, for example, can generate 1000+ named participants. This in

turn raised the question of validation and quality in multi-author listings in the outer fringes of the publishing universe if a "free for all" were to develop. I further noted for Wim that: *"One can imagine co-authorship of large multi-author papers being sold by some of the more unscrupulous occupants of this commercial space on a large scale, if this does not already happen. Therefore, what quality thresholds and checks would be set for full author listings, and how would malpractice be identified and weeded out?*

Clearly, there was much to consider here, but in the specifics of Dr Beigel's request, it would seem appropriate to complete the collaborative author list if technically possible to do so."

Wim observed that it was Scopus policy to capture all authors who were listed as an author or part of a collaboration group on the paper. There was no maximum in the number of authors being captured from the paper. In fact Scopus also includes so called **kilo-papers** (papers with more than 1000 authors) with all authors captured. He wrote that:
" In John's exemplar, I see that Scopus has captured 40 authors from the original paper along with [the ACTT-1 Study Group Members]. However, the individual members of the study group are not listed and do not form part of the metadata of the paper. ..The list of the ACTT-1 Study Group Members in the appendix may have been overlooked when processing the article.*
*nejm.org/doi/suppl/10.1056/NEJMoa2007764/suppl_file/nejmoa2007764_appendix.pdf

Peter Berkvens of Elsevier asked whether the statement '*The following study group members were all closely involved with the design, implementation, and oversight of the ACTT-1 study*' defined authorship. He pointed out the need to de-duplicate identities in some author lists. John Beigel replied that *"we are open to better ways to write it in the manuscript and in the appendix to clearly delineate this. If you have guidance or suggestions...."*

I noted the risk of abuse of this process if very loosely affiliated members of contributing units claim co-authorship in other large research projects. The lesson of this exchange seemed to be that SCOPUS should make clear that in future, only authors named in advance can be considered for a SCOPUS listing, and it was the responsibility of the lead authors and of the journal editors to enforce such policies, for which public education would be needed.

The SCOPUS CSAB Virtuall Meeting of October 19th – November 15th, 2021

Figure 19: The UN's sustainable development goals (see text)

The Covid pandemic rolled on through 2021, so we again adopted the model of four Zoom sessions for the Autumn 2021 meeting. In matters of policy, we discussed:

- Early on-boarding for selected Journals
- The Integrities of Trust
- Multi-language machine translation and its impact
- China Journal Analysis using Scopus data
- Societal challenges, and the UN's 17 sustainable development goals (Figure 19).
- Improvements to the SCOPUS search results system,
- New data architectures and the new search engine for SCOPUS.

Policies of Trust and the new SCOPUS Research Data Platform

Rob Schrauwen and Maxim Khan of Elsevier discussed the evolution of the SCOPUS Research Data Platform, whose purpose was to provide a current and reliable representation of the proprietary data in the SCOPUS system, including that which we had curated; and to make optimal use of network effects across organisations, publishers and funders, as illustrated in Figure 20.

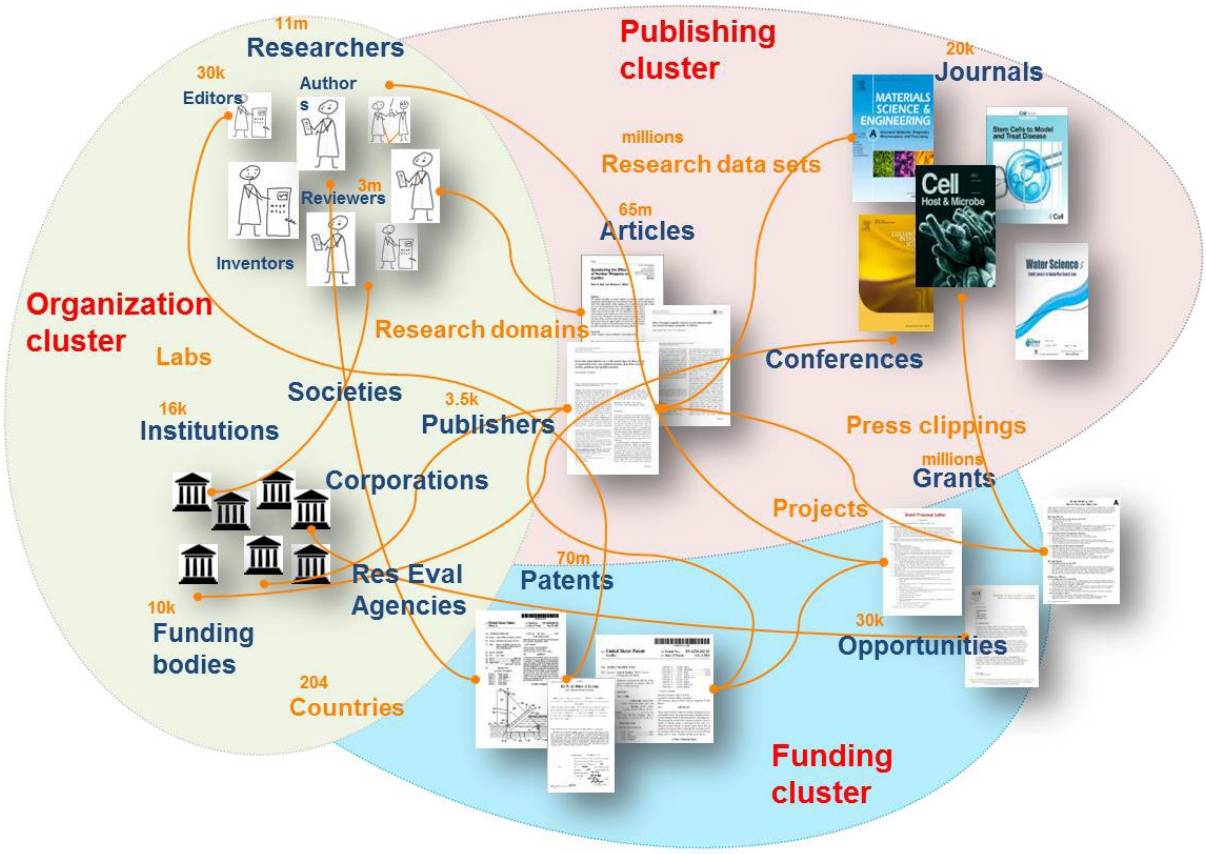


Figure 20: The network of entities in Academia for which SCOPUS data existed (see text)

They explored the concept of Entities, which include Works; “Containers” of Works; Persons; Organisations; Funding; Concepts (Subject Classifications); Research Indicators; Years; Equipment; Events (Conferences); Countries and Territories; and Publishers.

In respect of data on Entities within SCOPUS, we considered:

- How do we ensure our representation is trusted?
- Is there a principled approach to what is in or out?
- Is it of high quality, and does it have the right level of completeness?
- Do we have suitable research indicators and metrics?
- How to adapt Curation Policies for different user cases and data collections (such as unique national data collections, as for example In Japan and Egypt.

The Probabilistic Nature of Entity Data in the SCOPUS System

Rob and Max observed that our representation of information in SCOPUS is **Probabilistic**. Occurrences are matched with known registered entities (with certain accuracy) by the Person Entity Resolver Module. If no match is found, a new entity is registered, but only where there is credible evidence for it. This process demands attention to detail.

The registration process is necessarily incomplete, because SCOPUS only sees what is in the collected data, which is mostly derived from journal articles (Figure 21).

Figure 21: The reduction in Dimensionality during the journal data curation process.

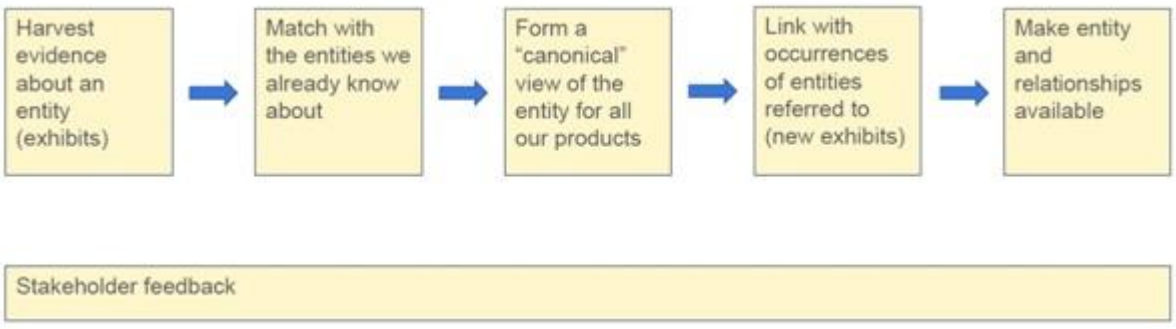


Figure 22: The process of the construction of entities in the SCOPUS database.

Rob and Max explained that the concept of the Elsevier Research Data Platform (RDS) was based upon the identification of data needs; the evaluation of the data strategy; the available sources, including their quality and usability; and their appropriateness of fit to the intended outcome. In their model, the framework comprised three layers (Figure 23):

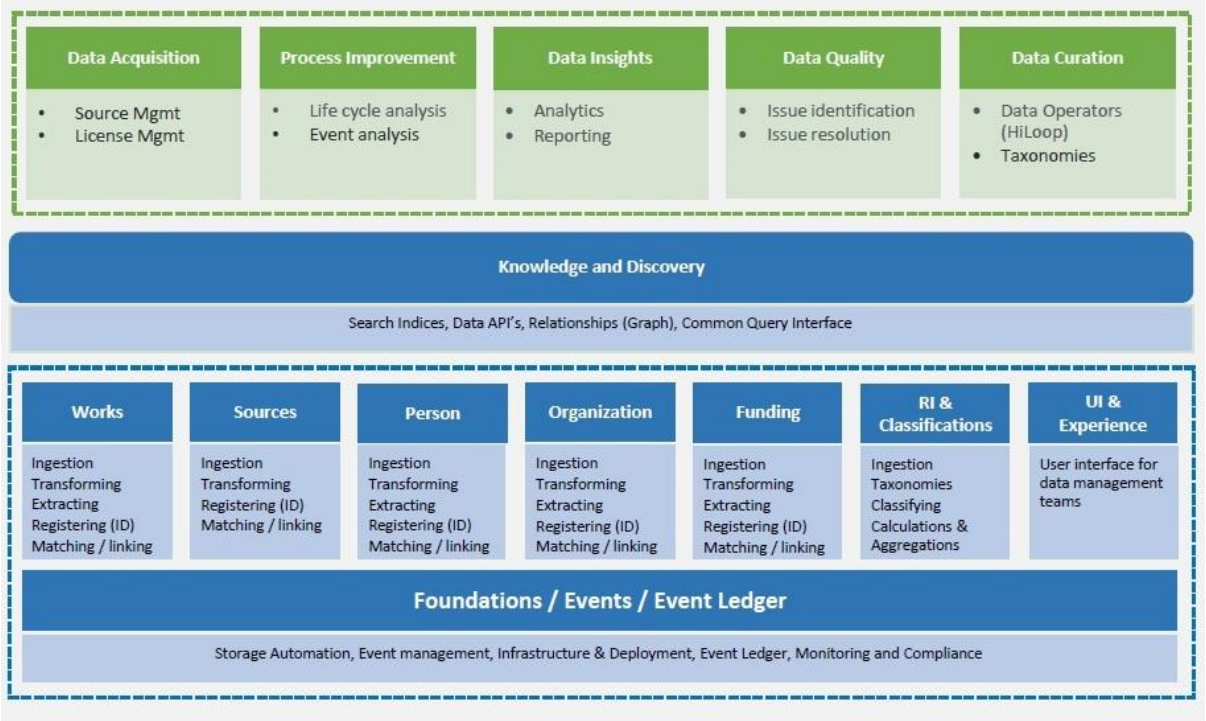


Figure 23: The framework of the Elsevier Research Data Platform. Key elements are:

- **The Data Layer**, of acquisition, improvement, insights, quality and Curation;
- **The Knowledge and Discovery Layer** of search indices, data APIs, and query interfaces;
- **The Foundation, Events and Events Ledger Layer** of works, sources, persons, and so on.

Collection	Use Cases
CSAB Journal Collection	Evaluation, Discovery
"Secondary" documents Collection	Discovery
Medline Journal Collection	Discovery
Conferences Collection	Evaluation, Discovery
Books & Book Chapters Collection	Evaluation, Discovery
Preprints Collection	Discovery
Patent Collection	Discovery
Organization profile Collection	Evaluation, Discovery

Figure 24: The primary collections of data within the SCOPUS system.

This information framework was in turn developed from the primary collections of data within the SCOPUS system, to whose policies and quality assurance the CSAB members were making a significant contribution (Figure 24).

An Operating Model for the Policies and Integrities of Trust

Figure 25: Tthe relationships of the Integrities of Trust (Courtesy of Elsevier)

Rob and Max then considered the relationships in the data collections, in which they asked:

Of Content & Data Policies:

- What content sources are selected (and which not)?
- Of those sources what is covered in the collection?
- Are we using a principled approach to selection?

Of Content & Data Quality:

- Is the current data quality sufficient to address the use case properly?
- When changes occur, how does this impact the data quality and outcomes?

Of Research Indicators:

- Were the definitions and values consistent across the ecosystem, and was there a clear explanation for any variation?.
- Do new indicators provide appropriate insights?

Of Authorities:

Is our data used by other organizations with the reputation to endorse it?

Of Data Privacy:

Is the user privacy and integrity properly addressed in the way how we reuse data and record the activity data of our users?

For users to have confidence and trust SCOPUS data, the Policy, Quality, Insights, Authority and Privacy are essential (Figure 23). For all activities on the Data Platform, the integrities need to be validated to demonstrate that they support and do not compromise Trust.

Practical Examples of the Importance of Data Integrities

Country Names provide a particular challenge where sovereignty is disputed, as in the governance of Crimea by Ukraine and Russia, and of Taiwan by (Mainland) China.

The Elsevier Integrities Team of three nominees meets on regular basis to discuss integrities and specific cases, and to make recommendations on how best to protect and build trust. The stakeholders who identify the needs include the managing editors of journals.

The guiding corporate principle is to play a neutral role and to provide the data in a normalised, structured way, up to the level of aggregation at the threshold of being undisputed. The data user then decides how they want to analyze or aggregate the data.

In primary publishing, it is left up to the author to decide what country name is being used, as long as it is technically correct, and the country name is validated to be an official, internationally recognised country to avoid the use of non-existing, fake country names.

For the implementation there needs to be a definition or overview of what are disputed regions (such as Taiwan, Crimea). Input is sought from the local teams on the use and wording of regions (such as 'Mainland China' or 'Hong Kong Special Administrative Region'); and on the likely impact and decisions on product functionality, such as search filters, or country level aggregation. Elsevier also seeks to be the purveyor of trust through openness.

The Policy on Self-citations

Self citations emerge from various sources, including Authors, Editors, Editorial Board Members, Peer Reviewers, Journals and Institutions. High self-citation ratios can be an indicator of citation malpractice, as for example in the exercise of undue influence by editors in demanding citations to their personal work. However, not all self citations are inappropriate, and they can be useful in highlighting new areas of research and recent

developments. We agreed that transparency in self-citations enables users to interpret the data and to use it as they see fit. The SCOPUS Radar system has subsequently evolved with the facility to identify and quantity patterns of self citation within SCOPUS data with considerable acumen and accuracy.

Policy on Modernisation of the Scopus Title Evaluation Platform

By 2020, it had become apparent that the STEP platform was in need of a major modernisation for a number of reasons:

- The legacy software and design was now 10 years old;
- The system was bespoke and lacked adaptability;
- The system had been conceived in a different era of trust;
- The processing of applicant journals was relatively slow and inefficient;
- Newer systems such as the SCOPUS Radar were not yet integrated into STEP;
- Our feedback to applicants would benefit from new approaches;
- Richer information flows were needed for journal review and evaluation purposes.

Specific technical problems included:

- Limited email and mailing functionality;
- Complicated data import routines;
- A dated look and feel to the interfaces;
- All routines were hard coded, and changes were expensive of time and resources;
- Development costs had escalated;
- Bibliometric data accuracy was not guaranteed and required manual checks;
- There were problems of stability and performance with file uploads;
- The system was not compatible with other Elsevier systems;
- The sustainability of the supply chain was uncertain.

The case for modernisation and investment was clear and agreed, and a plan was set in motion to implement the necessary changes.

The Impact of Machine Translation on the English Language Policy

Since the creation of SCOPUS, English has been mandated for titles and abstracts in SCOPUS. English has a range of fault tolerant features which have permitted it to evolve into the world's primary second language, which I had explored in a paper for Russian Science Editor and Publisher in 2017. (doi:10.24069/2542-0267-2017-1-14-18)

However, a number of factors prompted a rethink of this policy, in that:

- The policy was a constraint on journals in other languages and non-latin scripts;
- Machine based translation and speech recognition systems were improving rapidly
- The technologies applied to translations to and from English, or from any language to any language,or from any script to any script.

Much had happened in the worlds of technology and publishing since Douglas Adams described the Babel Fish as a Universal Translator in the Hitchhikers Guide to the Galaxy in 1978. Machine/AI inter-language translation tools such as Google and Apple Translate had advanced dramatically, to the point at which we could ask whether content for the global reader will soon become "source language agnostic"

Moreover, most journals and many books have moved from analogue to digital formats. It should therefore be possible to put machine translation systems both at the publishing point and / or at the reader point to minimise the barriers, costs and challenges of translation on an article by article basis. I therefore wondered:

- How could such systems be implemented?
- How might their use influence the STEP decision rules and processes?
- What protocols would be applied to ensure reasonably accurate translation?
- What kitemarks would apply to validate academic translation systems?
- What "Human in the Loop" principles would need to be applied?
- How would error correction be captured and fed back into further refinement?

I appreciated that machine translation may not yet handle the subtleties of Arts and Humanities content, but the language of Medicine is fairly standard globally, and it might make a good test case for a broader approach.

Elsevier was already making some progress in automated translation, including:

- Automated translation for a French Module in Embase for pharmaco-vigilance.
- Use of non-Roman script in linking and connecting data.
- A pilot with Chinese journals with references in Chinese.

Our Discussions on Russian <-> English Machine Translation

Natalya Popova of the Russian Board and a Senior Researcher at the Centre for Theoretical Linguistics and Academic Communication of the Ural Branch of the Russian Academy of Sciences, noted that:

"I have been working in the translation industry (mainly Russian into English) for several years now. AI is increasingly becoming our right hand. Today, when more and more translators rely on their Computer-Assisted Translation (CAT) tools, we mostly talk about the importance of editing rather than translation competences.

I have noticed that AI works almost perfectly in some fields, but only if the source texts are of sufficient quality. In Russian-English translation, the "machine-friendly fields include medicine, maths and IT, where the terminology is quite straightforward for Russian-speaking authors. However, AI can only provide the very basic translation when it comes to discipline, which have developed in a degree of professional isolation during the USSR period – including engineering, physics, humanities and the social sciences. These disciplines have developed their own terminologies and phrasing in Russian, which are quite different from those used in English.

Another important issue is the quality of source texts. Many Russian authors still write long, incomprehensible sentences, and their machine translations are not of publishable quality. The linguistic quality of texts may become another marker of the journal's level."

I have observed that we have the same problem of long, incomprehensible sentences in English academic writing. There are many authors who still believe that such writing is a sign of higher intelligence rather than a disorganised brain! The use of clear, simple, short, active sentence and paragraph structure is a fundamental skill of communication in any language.

I had not previously fully considered just how important is clarity and simplicity in the collective shape and flow of words in any language for efficient AI assisted translation, rather than just the mechanical translation of individual words and phrases.

Ashok Raina (Subject Chair for Mathematics) recalled his only attempt to use the services of a professional translator for a paper in Russian. He wrote that: *"The translation service was run by the Dept. of Atomic Energy, but it was apparent that the translators were more capable of translating Pushkin than mathematics! I have not tried Google translate on any mathematics paper but I wonder how it manages to distinguish the mathematician's use of terms like bundle, sheaf, germ, field, group, surgery from their other uses?"*

Natalya replied by highlighting how far AI has developed since Ashok's experience. She noted that *" By my first university degree, I am a mathematician, so I understand perfectly well what you mean.... computer-assisted translation (CAT) tools, which are based on Google, Yandex (The Russian version of Google) and other neural machine services have the major advantage of translation memory, which is accumulating the more texts you translate... Such systems make the whole translation process very fast and rather efficient. Nevertheless, the role of a human is still indispensable! Machines still make some funny and awkward mistakes :)"*

A letter in The Times (of London) on 31st May 2021 added spice to the conversation on the importance of "human in the loop" translation. The correspondent wrote:
"Sir, Gerard Baker in his article on China (May 28) uses the menacing Nikita Khrushchev quote: "We will bury you." But did Khrushchev actually say this? A different translation is merely: "We will attend your funeral." This can be paraphrased as "Our Communist system will outlive your capitalist system." Khrushchev was wrong in the long run, but the meaning of his words is significantly softened and rendered less belligerent, if we accept this alternate translation instead of a biased Cold War cliché."

Natalya observed that "*This quote of Khrushchev is very famous, but it is not an example of scientific mistranslation. Khrushchev used a Russian idiomatic expression, and the interpreter failed to find an adequate equivalent on the spot. However, in some disciplines (eg*

economics), researchers also use colloquial expressions. I once faced a dilemma when translating an Economics paper, which used Lenin's famous saying 'Верхи не могут, а низы не хотят' to describe the situation preceding a revolution. If translated literally it sounds nonsense "The upper classes do not want, but the lower classes cannot». It took me some time to come up with something like «The rulers are unable and the ruled ones are unwilling to go on as before»".

Our Paper on Machine Translation in Russian Science Editor

Natalya and I summarised our deliberations in a joint paper for Olga Kirillova's journal Russian Science Editor (Rew and Popova 2021). We observed that:

"Clear translation remains a major challenge to better communication and understanding of the international academic literature, despite advances in Machine Translation (MT).

Automatic translation systems which captured the detail and the sense of any manuscript in any language for a reader from any other linguistic background would find global applications... including the facility to skim and scan large amounts of information in foreign languages, and to act as digital dictionaries, thesauri and encyclopaedias. Word- to-word and phrase-to-phrase translation between many languages and scripts is now well advanced.

The availability of modern machine translation has therefore changed the work of specialist scientific translators, placing greater emphasis on more advanced text and sense editing skills. However, machine translation is still challenged by the nuances of language and culture from one society to another, particularly in the freestyle literature of the arts and humanities.

Scientific papers are generally much more structured, but the quality of machine translation still largely depends on the quality of the source text. This varies considerably between different scientific disciplines and from one author to another.

The most advanced translation systems are making steady progress. It is timely to revisit traditional training programmes in the field of written translation to focus on the

development of higher level research competencies, such as terminology search, and so to make best use of evolving machine translation technologies.

More widely, we consider that there is a challenge across the higher education systems in all countries to develop a simple, clear and consistent "international" writing style to assist fast, reliable and low cost machine translation and hence to advance mutual understanding across the global scientific literature".

The Settled Policy for English Language Use in SCOPUS in 2021

Wim Meester summarised our deliberations from the SCOPUS policy perspective thus:

"The message is that word to word translation (eg for restaurant menu ordering) is well established, but machine translation is still too immature for reliable and accurate computer and AI-mediated sense and context translation from <Language A> to <Language Z>...

The current SCOPUS policy of using English as the common medium for titles and abstracts is therefore defensible as the best currently available option for international academic publication and communication.

However, this is a rapidly advancing field, and we should keep the subject under review. The point is also clear that high quality, clear, concise, standardised and focussed scientific writing in any language will made optimal machine translation much easier to achieve in future, even if it makes individual texts more boring to read!

The current SCOPUS policy of using English as the common medium for titles and abstracts is therefore defensible as the best currently available option for international academic publication and communication.

Five years on, there has been an explosion in the functionality in the machine-enhanced interpretation of language, with many generative AI products in public use, and the major search engines now offer translation on demand for many languages. The questions arise as to whether:

1. Technology has moved on far enough with Gen AI to justify another look at the issue of machine translation (MT); and

2. MT has passed an inflexion point at which it could reliably be put in front of a publisher's outputs, as Google Translate does for web content, so that content in any of a wide selection of languages could be put directly in front of readers in a "good enough" translation.

3. The presentation of SCOPUS language policies and their practical purpose to non-English language audiences should now be reviewed in late 2025.

Citation Metrics, Altmetrics and the Actual Usage of Published Papers

I have long been concerned at how limited are citation metrics as a measure of societal impact. In looking back through my own publication history, I was very struck at how remote bibliometric measures are from practical reality. For example,

- A 2003 textbook chapter on Cell Imaging Techniques has received 3 citations but 1700 downloads, according to Springer Link
- A 2019 paper on the future Electronic Patient Record in a trusted BMJ offshoot Journal has so far had three citations, but more than 4000 full article downloads.
- This paper also attracted a burst of altmetric comment in the first few days of release, but nothing subsequently, suggesting to me that altmetrics scores may largely be a transient phenomenon for recently published papers.

Trustworthy Article Download numbers would therefore offer a much stronger index of practical interest e of content than of citation activity, but download statistics would be highly susceptible to manipulation and therefore generally untrustworthy.

Peter Brimblecombe (Environmental Sciences) added that:

I have been recently doing citation analyses of the top cited papers in the specialist field of, heritage modelling. Most citations to even these very important papers are trivial mentions in the introduction and are in-substantive. It made me quite gloomy to have been forced to see citations as such a strange measure of a published paper's success.

Downloads, as with citations, are often manipulated, so they need to be analysed carefully. It is useful to ascertain that they come from different IP addresses, different countries and are not time stamped at the same time each day. However, they are valuable as they can reflect a kind of immediacy and use, not captured in citations. I also find that readership on ResearchGate gives clues to the use of a paper.

When I was a PhD student in the early 1970s, citations were used in a very different way. My supervisor explained to me that reference lists track a topic back in time and citations track the topic forward. I found this amazing and useful, but that was all done on paper and no one had then thought of the Impact Factor or H Index."

Wim Meester noted that Download Metrics give a good indication of the popularity of an article, as evidenced by a paper by Ding et al in *Scientometrics*, in which the authors studied 7552 articles which had been published in *The Lancet* from January 2007 to October 2017. They had noted that "most articles are downloaded more at the early stage after publication, whereas the citations mainly appear after a period of time, making the slope trend of downloads and citations inconsistent throughout the citation life cycle.

Wim also noted that with the recent acquisition of Plum Metrics, Scopus also provides metrics on captures (bookmarks, readers in Mendeley, etc.), usage (clicks, downloads, views), mentions (blog posts, news media, etc.), social media and citations (including citations from local citation databases). This gives a more broad impression of the impact the article makes. As of 2021, Scopus also provided metrics on the view counts of an article.

While bibliometric measurements continue to rule the roost in quality assurance and performance evaluation, the search for meaningful and trustworthy supplementary information sources continues.

In Summary

Despite the challenges of the Covid pandemic through 2020-2021, this was a very productive period for discussion and progress in many aspects of policy making for the SCOPUS Content Selection Advisory Board.

In the next essay, I consider the further work of the board from 2022 to 2025.

Acknowledgements

I am grateful to many colleagues for discussions and insights in this subject during my tenure as the Subject Chair for Medicine for the SCOPUS CSAB over two decades.
I am also grateful to Professor Julie Cullen of the University of Southampton for proofreading and comments during the preparation of the manuscript.
The recollections and opinions in this essay are my own. They should not be construed as representing the corporate views, policies or intent of Elsevier BV.

Postscript: Depatures from the Board during the Covid Pandemic

The Departure of Professor Evan Bieske (Chemistry)

We were very sorry to note Professor Evan Bieske's retirement in early 2021 from the Board to pursue other interests. Evan had been a dynamic and entertaining contributor throughout my time on the board since 2009, and an energetic debater.

The Death of Professor Peter Stambrook

On June 4th 2020, we learned of the forthcoming retirement through ill health of Peter Stambrook, who had been a stalwart of the Board since the launch of SCOPUS in 2004 as the Chair for Genetics, Pharmacology, Toxicology, Pharmaceutics, Biochemistry, Molecular Biology and Neurosciences.

Peter had become a particular friend and mentor on the board, and his aggressive malignant acral melanoma was soon to lead to his cruel and untimely death 1st October 2020.
(see legacy.com/us/obituaries/cincinnati/name/peter-stambrook-obituary?id=10261809)

References

Cortegiani A et al: bioRxiv 2020.03.26.007435; doi.org/10.1101/2020.03.26.007435
And subsequently published in *F1000Research* doi: 10.12688/f1000research.23847.1

Macháček V, Srholec M. Predatory publishing in Scopus: evidence on cross-country differences. Scientometrics. 2021;126(3):1897-1921. doi: 10.1007/s11192-020-03852-4.

Macdonald, S. (2023). The gaming of citation and authorship in academic journals: a warning from medicine. *Social Science Information*, *61*(4), 457-480. https://doi.org/10.1177/05390184221142218 (Original work published 2022)

Macdonald, S. https://blogs.lse.ac.uk/impactofsocialsciences/2025/02/17/does-authorship-mean-anything-when-academic-papers-are-simply-citable-tokens/

Ding, Y., Dong, X., Bu, Y. *et al.* Revisiting the relationship between downloads and citations: a perspective from papers with different citation patterns in the case of the *Lancet*. *Scientometrics* **126**, 7609–7621 (2021). doi.org/10.1007/s11192-021-04099-3

Selected References to my Other Papers in this Series on Academic Publishing,

Rew DA. - The development of high impact national and regional journals in medicine and the health sciences Russian Science Editor and Publisher Issue 2020; 5(2): 113–122 doi 10.24069/2542-0267-2020-2-113-122

Rew DA. The role of English in International Publishing. Russian Science Editor and Publisher 2017; 2 (1): 14-18. DOI:10.24069/2542-0267-2017-1-14-18

Rew, DA., Popova, N. The challenges of machine translation of academic publications. Russian Science Editor and Publisher, [S.l.], v. 6, n. 2, p. 104-112, nov. 2021. ISSN 2541-8122. https://www.scieditor.ru/jour/article/view/162/131 doi:10.24069/SEP-21-01.

These papers have been published on the University of Southampton ePrints server

Rew DA and Lehman M. An Audit of Correlation of the MEDLINE and Scopus Evaluation Systems for the Quality Assurance of Academic Healthcare Journals https://eprints.soton.ac.uk/502915/

Rew DA. Preparation for Editorship of the EJSO, 1996-2002: https://eprints.soton.ac.uk/502916

Rew DA. Editorship of the European Journal of Surgical Oncology (EJSO), 2003 to 2009 https://eprints.soton.ac.uk/501604

Rew DA. The Development of the SCOPUS Content Selection Advisory Board, 2003-2009 https://eprints.soton.ac.uk/507668/

Rew DA. "The Complex Universe of Citation Data for Bibliometric Systems" https://eprints.soton.ac.uk/502914/

Rew DA. "The Case for the Evolution of the All Sciences Journal Classification (ASJC) System" https://eprints.soton.ac.uk/504370

Rew DA. "Trust, quality assurance and the classification of academic publishers" https://eprints.soton.ac.uk/504371/

Rew, DA (2025) *The Evolution of the SCOPUS Title Evaluation Platform (STEP), 2010 to 2025* 39pp. https://eprints.soton.ac.uk/507702/

Rew, DA (2025) *The development of the SCOPUS Content Selection Advisory Board, 2003-2009* https://eprints.soton.ac.uk/507668/

Rew DA (2026). The development of policy for the SCOPUS Citation System Part 1: the work of the Content Selection Advisory Board 2010-2011 https://eprints.soton.ac.uk/508986 /

Rew DA (2026). The development of policy for the SCOPUS Citation System Part 1: the work of the Content Selection Advisory Board 2012-2016 (soton eprint server in press)

Rew DA (2026). The development of policy for the SCOPUS Citation System Part 1: the work of the Content Selection Advisory Board 2017-2019 (soton eprint server in press)