

**Modernising the Academic Subjects Journal Classification (ASJC) System –  
A Discussion Paper**

**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.

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

Open source Preprint for publication on the ePrint Server, University of Southampton

28th February 2025

Correspondence to  
dr1@soton.ac.uk

**Key words**

Academic subject classification; All Sciences Journal Classification; Scopus; Web of Science;  
Medline; Ulrich's Periodicals; Article level classification

## **Abstract**

**Background:** The classification of documents, articles and journals provides structure, logic and integrity to a large and proliferating ecosystem of academic and research information. Reliable classification systems are of value to many users, including publishers, institutions and authors who are looking for the most appropriate repository for their publishable work.

The All Sciences Journal Classification System (ASJC) is the best known of such systems. It was developed more than 20 years ago by a team at Elsevier Science. The academic corpus now includes many subject areas outwith the original design of the ASJC, as in the Arts and Humanities. The Internet has also fostered a substantial growth of non-traditional publishers and publication methods, and a proliferation of multi-subject and multidisciplinary journals.

**Methods:** 15 years of observation of the use of the ASJC in the quality assurance of journals has provided subjective evidence of the limitations of the original ASJ Classification. The SCOPUS database was therefore used to analyse the distribution of journals across the present classification, and to match a machine generated, article-based classification against the ASJC, based upon the titles and abstracts.

**Results:** The analyses confirmed the very uneven distribution of journals by subject across the ASJC, and the lack of correlation between the calculated article-based classification and the allocation of journals using the existing subject based classification. Multidisciplinary journals pose particular challenges to journal classification systems.

**Conclusions:** The arguments for the modernisation of the ASJC are balanced against the development and introduction of article and author based classification systems. There is an increasing role for machine based and AI assisted tools to automate human expert based methods of classification. However, so long as journals remain a key element in knowledge dissemination, there is a strong case to be made for a modernised and regularly updated ASJC based all sources classification system for academic publications.

## **Introduction**

It is natural to seek to classify complex information systems to simplify their comprehension and navigation. The corpus of human knowledge which is represented by the publication of academic papers and articles in journals, Theses; academic books, textbooks and book series; patents; conference proceedings; preprints and policy papers at Government and Institutional level from many countries and in many languages, is one such complex ecosystem.

Many researchers and academics in the STEAM subjects (Science, Technology, Engineering, Arts, and Mathematics) continue to use conventional journals and Conference Proceedings to promote their outputs, but many other career academics and researchers measure their outputs and impacts through other formats. There is also a growing, informed and influential ecosystem of high quality academic and educational content on non-traditional platforms, including social media and academic blogs.

The vast ecosystem of academic publishing spans some 300,000 serial publications across all scholarly disciplines, as recorded in Ulrich's Periodicals Directory (Proquest LLC). Fewer than 30,000 active titles are listed as academic journals in the principal quality assured citation systems, SCOPUS and Web of Science. Many journals still exist outside the mainstream of quality assessment systems. Inactive and secondary journal sources may comprise a further 25,000 journals in SCOPUS.

SCOPUS listed Journals contain some 100 million individual academic papers. Journals are living instruments in a Darwinian knowledge ecosystem. They are born and evolve. The most successful thrive over generations, but many others fade away.

The All Sciences Journal Classification System (ASJC) is generally accepted as the primary global classification system. It was developed more than 20 years ago by a team at Elsevier Science. For the purposes of this paper, I will use the descriptor "Journals" flexibly to include other academic periodical publication formats..

### **The characteristics of a modern and logical academic subject classification system**

A logical academic subject classification system should be contemporary, adaptable and rationally structured. Given the continual evolution of knowledge in all subject areas, including the Arts and Humanities, it must be maintained and regularly updated as new disciplines and topics emerge, as for example, around Artificial Intelligence.

An ideal such system would allow the clear classification of all journals and other academic publishing vehicles into a logical and comprehensive subject framework, such that any journal could be cleanly pigeon-holed according to its title, aims and correlated content (Figure 1). However, the classification of many academic journals does not fit readily into an academic subject framework.

Some journals publish on speciality subjects which are clearly defined by their title, aims and scope, as for example, the European Journal of Surgical Oncology (EJSO) and the BJS, formerly the British Journal of Surgery. Other journals have changed from their original briefs, for example in taking papers from authors on a global basis rather than from a defined geographic area.

However, many other journals are multidisciplinary in content, and cannot be so easily pigeon-holed. They may provide wide intra-disciplinary subject coverage, as for example the British Medical Journal, the New England Journal of Medicine, The Lancet, and Nature Medicine. These may be described as Subject-Specific Multidisciplinary Journals.

Across subject areas, other journals may take papers from any subject area, as for example Nature, Science, PNAS, PLOS1, and the Elsevier Journal Heliyon. These journals may be described as General Multidisciplinary Journals.

Moreover, many journals have drifted from their original aims and scope, or have set their titles, aims and scope so vaguely, broadly or generically as to defy any simple classification or correlation between title, aims and content.

It is therefore clear that a historic classification system which correlates journals and their titles with their content is no longer sufficient for practical purposes. It is also clear that the generic “multidisciplinary” classification needs much more rigorous analysis and subdivision. At present, coders and sorters of journals for various bibliometric systems are obliged all too frequently to defer to the “multidisciplinary” and “miscellaneous” categories in the ASJC in the absence of sufficient granularity and dimensionality in the present version of the ASJC .

The question therefore arises as to whether it is possible to evolve and adapt the existing ASJC and its counterparts to be fit for modern purposes, or whether a wholly new approach is required. A contemporary, reliable and workable classification system for Journals and other academic outputs may yet retain an important role for the academic publishing and bibliometric ecosystem for many reasons, which include:

- the formulation of policies and investment strategies for Institutional, Corporate, National and International organisations
- the clearer definition and detection of malpractice in Academic Publishing
- the direction of authors who are looking for a suitable home for their manuscripts.
- the support of researchers who are looking accurately and consistently to describe their academic productivity.

### **The All Science Journal Classification (ASJC) scheme**

A Google Search of Journal Classification systems highlights the All Science Journal Classification (ASJC) scheme (Figure 1) as the de facto primary Journal Classification system in common use.

As originally created, the ASJC is characterised by four high level subject areas (super groups), each of which has a number of key subject area sub-classifications (Figure 1). Each subject area is allocated a four digit code. For example,

Life Sciences subjects are allocated 1100 codes,

Arts and Humanities are allocated 1200 codes

Biochemistry, Genetics and Molecular Biology allocated 1300 codes

Subject Area	Subject Area Classifications
Physical Sciences	Chemical Engineering Chemistry Computer Science Earth and Planetary Sciences Energy Engineering Environmental Science Material Science Mathematics Physics and Astronomy Multidisciplinary
Health Sciences	Medicine Nursing Veterinary Dentistry Health Professions Multidisciplinary
Social Sciences	Arts and Humanities Business, Management and Accounting Decision Sciences Economics, Econometrics and Finance Psychology Social Sciences Multidisciplinary
Life Sciences	Agricultural and Biological Sciences Biochemistry, Genetics and Molecular Biology Immunology and Microbiology Neuroscience Pharmacology, Toxicology and Pharmaceutics Multidisciplinary

Figure 1. The four high level subject areas (supergroups) and the subsidiary high level subject classifications within the ASJC.

The Medicine classification within the Health Sciences super-group has been allocated the 27\*\* number, allowing for 100 further sub-classifications (2700 to 2799). Some 50 of these numbers are as yet un-allocated, leaving scope for greater detail in the sub-classification. (see <https://scientificresearch.in/asjc-all-science-journal-classification-codes/>.)

Within each of the ASJC “supergroups” lies a further layer of sub-classifications, to each of which are allocated journals according to the best judgement of the relevant bibliometric classification team. They will take into account factors such as the title of the journal, the aims and scope, and the subject mix of the individual articles in each journal.

The detailed origins, history, and maintenance methodology of the ASJC are uncertain. It is used particularly by Elsevier coders when a serial title is set up for Scopus coverage, and in

various forms by other organisations for classification purposes in the absence of a globally agreed and standardised system.

The widely used Quacquarelli Symonds (QS) ranking system evaluates institutions across five broad faculty areas and 55 subject areas, which are based upon ASJC codes. The Clarivate web site reports that in the Web of Science, categories are assigned at the journal level to one of 250 categories and that a journal may have up to six assigned categories to it, including “a journal's categorisation in other bibliographic databases” (1). The WoS Subject category web page shows this list of 250 subject headings by alphabetical order from Acoustics to Zoology, but it offers no other sub-classification or granularity. It has clear parallels to ASJC, but to an indeterminate degree in the absence of further information.

#### **Published commentary on the limitations of the ASJC system**

Qi Wang and Ludo Waltman (2016) (2) studied the accuracy of the journal classification systems of Web of Science and Scopus, with particular reference to the field of Library and Information Science. They noted weaknesses in the classification systems, in that some journals that had weak connections with their assigned categories, while other journals were not assigned to categories with which they have strong connections.

The issue of poor correlation of SCOPUS content with the ASJC was also forcefully highlighted in a blog post on Retraction Watch by Aleksandar Stević, under the title “Scopus is broken – just look at its literature category” (3):

These and similar observations prompted a series of studies and discussions with the SCOPUS and wider Elsevier Classification and Analytics teams around options for the evolution of the ASJC, vis-a-vis newer forms of article classification. The preliminary outcomes from these discussions are shared in this essay.

## Methods and Findings

In order to understand the current status of the ASJC in terms of content of the SCOPUS data base as categorised by the ASJC codes, Dr Rob Schrauwen and colleagues at Elsevier undertook a series of analyses the SCOPUS data set. He reports (personal communication) that:

*“Elsevier’s data science team had created a classifier which was designed for and trained on a corpus of grant award notifications. These are documents which we acquire and harvest from funding bodies across the globe, detailing grants given to research. In our knowledge graph we connect these to researchers, organizations and research output, and the subject classification which used the same terms as the Scopus corpus helped with this process. A propriety classification model, based on machine learning and widespread classifier techniques, was created.*

*The need to assign classifications for articles was motivated by two main reasons.*

- The first is that more and more use cases look at the articles themselves, and therefore classifying based on the journal’s subjects does not correctly represent the subjects in the corpus of articles.*
- The second is that increasingly, Elsevier combines the Scopus corpus with non-Scopus material, sometimes incorporating many hundreds of journals, whose subject classifications we don’t have and don’t need.*

*Although the academic articles differ from grant documents, there was reasonable success in extending and tuning the algorithm for research output. The following figures show results of an initial version of this model, but formal development has not been completed.*

*There are inherent complexities in applying the ASJC classification to articles. Multi-disciplinarity is a feature of a journal and not of an article. Hence, not all classification codes apply to article classifications, and a separate “ASAC” subclassification scheme has been proposed. In consequence, the design of an algorithm to aggregate this to journal level is nontrivial. For the figures supplied, a simple aggregation was used based on frequency. This model is not suitable for commercial production use but it provides useful insight into the accuracy of journal subject classification.”*

The distribution of numbers of journals by the current classification is shown in the histogram in Figure 2. The chart highlights the substantial disparity in numbers from one category to another, with a particular peak in the 2700 (General Medicine) Category. This very large peak reflects the lack of granularity in the ASJC, in consequence of which a very large number of journals have been allocated the holding code for want of a more granular Medical subject classification across which to reallocate these journals more closely to their specific clinical subjects.

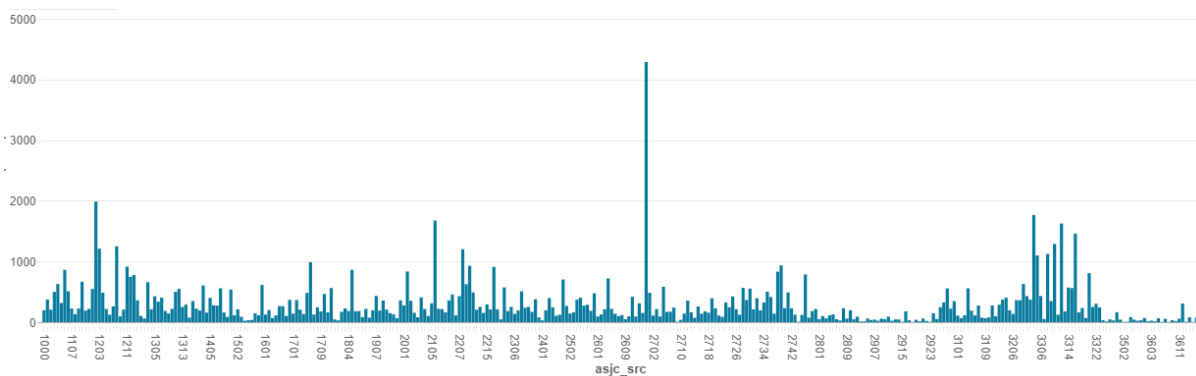


Figure 2: This displays the Number of academic journals per ASJC category. The X axis displays each of 334 categories in 27 broad subject areas (in thousands). The Y axis displays the number of journals in each ASJC category.

(see text. Figure courtesy of Dr Rob Schrauwen, May 2024)

Figure 3 similarly displays the number of individual documents which are listed in SCOPUS as they are associated with the ASJC codes through the journals in which they are published. We refer to the underlying document allocations as “aspirational”, in that they reflect the documents which journals have published through their aspirational aims and scope. The largest number of documents by far remain associated with the 2700 General Medicine. On this graph, the Y axis displays the numbers in millions of documents.

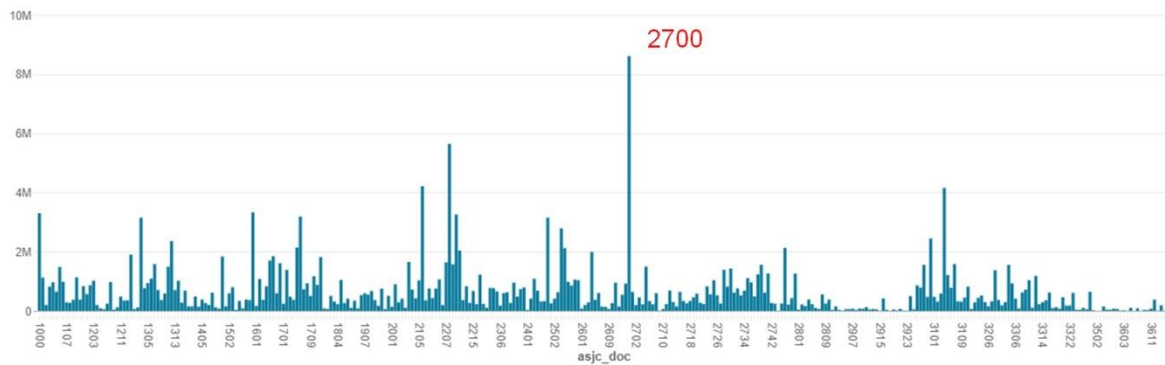


Figure 3. This figure displays the number of documents per ASJC code, as originally allocated to the journal. (see text). Figure courtesy of Dr Robert Schrauwen May 2024)

In an experimental study using a proprietary algorithm, Dr Rob Schrauwen and colleagues at Elsevier (Personal Communication, May 2024) analysed every document in SCOPUS and applied the machine generated allocation of the article to the most appropriate subject category, as from the “observed” nature of the article, rather than that which had been dictated by the journal in which it had been published (the “aspirational” allocation). The redistribution of articles was striking. For example, articles from journals in the Medicine (code 2700) were widely redistributed across the medical specialities for which specific codes exist, and to Code 2208 (Electrical and Electronic Engineering) as shown in Figure 4.

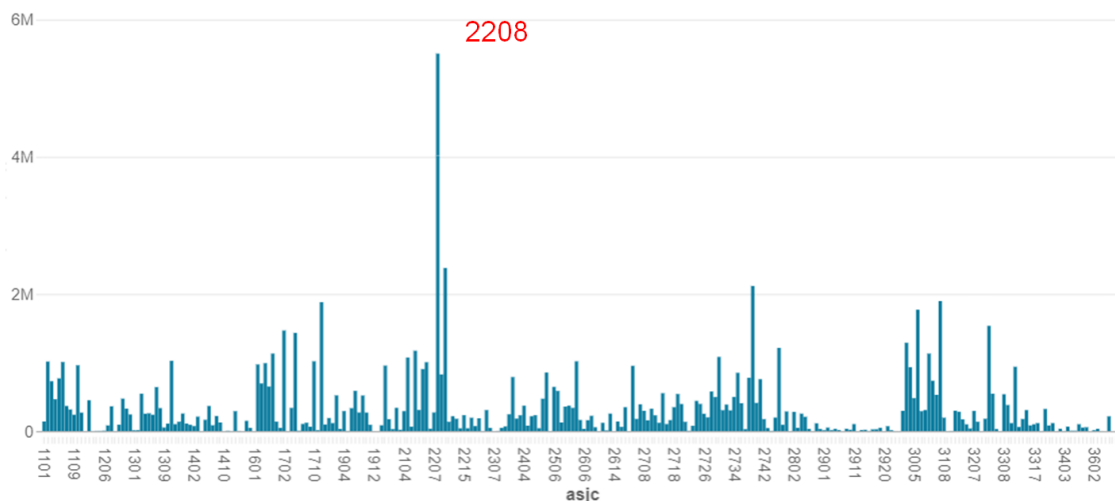


Figure 4. This figure displays the redistribution of documents in SCOPUS from existing subject categories to observed subject categories to after application of an experimental proprietary algorithm. The Y axis displays the number of articles. The X axis displays the ASJC subject codes. (see text)



Figure 5. This illustrates the distribution of numbers of articles by ASJC code in the physical sciences. The one significant outlier is the 2201 (Engineering, Miscellaneous) code, with negligible numbers. This demonstrates that at least at the macro level, the Physical Sciences subject codes are sufficiently granular to accommodate all of the reclassified documents.

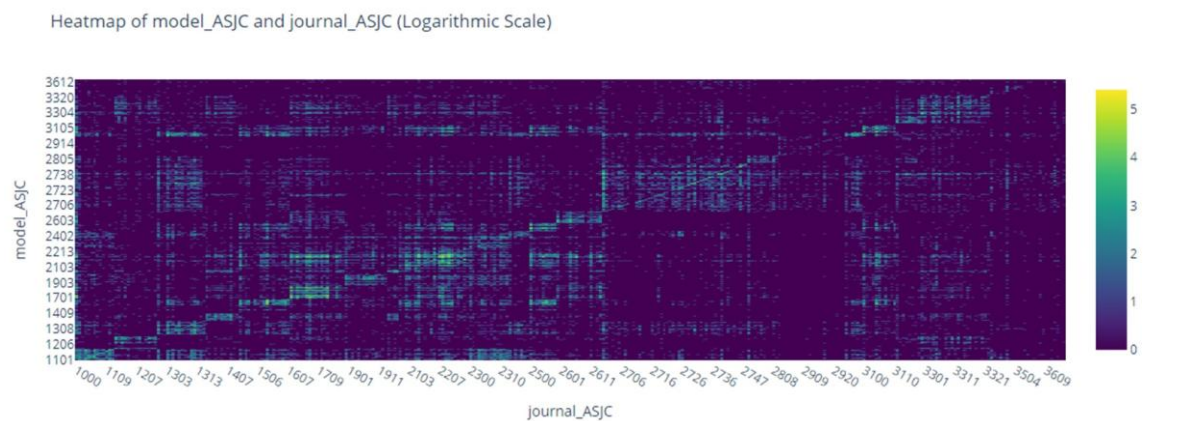


Figure 6. A heatmap of an article-based model of ASJC subject allocations, (Y axis), plotted against the journal-based ASJC classification in which the article was published (see text) Image courtesy of Dr R. Schrauwen)

This movement of articles and documents between subject areas is also illustrated in the Heat Map in Figure 6. In this figure, the algorithmic calculation of the most appropriate category (observed) for each article has been plotted against the category (aspirational) to which its parent journal had originally been allocated. If there were complete concordance

between the calculated (new model) and pre-allocated parent journal classification, there would be a diagonal linear relationship between the two groups.

In practice, while this relationship is discernible for many articles, it is also clear that there is a substantial difference between the new model “best fit” allocation of the article and the old subject allocation of the journal in which each article was published. The brightness on the colour scale reflects the number of articles which have been reallocated to each pixel block.

This article based classification model may need further refinement. However, this first pass analysis highlights the challenges of how best to allocate both articles and journals to the most appropriate ASJC subject classification code (Figure 6).

These studies have also demonstrated that:

a. Journals with SCOPUS listings not infrequently appear to have been allocated to seemingly inappropriate subject categories. This observation is also evidenced by the SCIMAGO journal performance listings. SCIMAGO draws journal lists from SCOPUS according to the ASJC, and republishes a range of performance metrics, by arrangement with Elsevier.

This mismatch is illustrated in Figure 7, which is a contemporary screenshot of the Anatomy category, which was chosen at random from the Medicine subject fields. At first glance, Journal No 1 (American Journal of Surgical Pathology) would seemingly belong in the Pathology category, as would Journal 5 (Journal of Histochemistry and Cytochemistry), while Journal 2 (Human Brain Mapping) belongs in Neurosciences or Radiology.

Journals are allocated an ASJC subject classification on accrual to the SCOPUS system by trained but non-subject specialist staff, on the basis of the journal title, aims and scope, rather than upon a quantitative analysis of the article types within. Once allocated, journals are very rarely if ever reallocated to a different code, so they carry any original and substantial allocation errors indefinitely.

Medicine		Anatomy	All regions / countries				
<input type="checkbox"/> Only Open Access Journals <input type="checkbox"/> Only SciELO Journals <input type="checkbox"/> Only WoS Journals							
Title	Type	↓ SJR	H index	Total Docs. (2023)	Total Docs. (3years)	Total Refs. (2023)	
1 American Journal of Surgical Pathology	journal	1.723 Q1	240	159	600	5111	
2 Human Brain Mapping	journal	1.626 Q1	217	438	1138	31422	
3 Advances in Anatomic Pathology	journal	1.600 Q1	90	50	115	3466	
4 Anatomical Sciences Education	journal	1.570 Q1	71	111	271	6930	
5 Journal of Histochemistry and Cytochemistry	journal	1.177 Q1	137	44	164	1441	

Figure 7; A screenshot of the SCIMAGO journallistings for Anatomy, as of Feb 2025 (see text)

### b. Journal Classification and the (Miscellaneous) Categories

The lack of sufficient granularity in the ASJC is further emphasised by the use of the (Miscellaneous) category in Journal allocation in each major subject area. This category has been created as a wrapper for journals which fit broadly within a super-group but which do not fit clearly into an existing specific subject classification.

The Scimago website lists the journals and ranking of all journals which are published by SCOPUS and Web of Science by a number of data points and characteristics such as the country of origin, and it is a very useful reference resource (see

<https://www.scimagojr.com/journalrank.php>).

### **The lack of granularity in the ASJC**

The SCIMAGO journal ranking model is also very useful for visualising the existing problems with journal classification. For example, it somewhat unhelpfully contains more than 4000 active and inactive journals in the Medicine (2700) category as of December 2024 (Figure 2).

This is because the ASJC schema is insufficiently granular in descriptive terms to accommodate many of the obviously “Medical” journals, even though there are still around 50 four digit codes in the 27\*\* group which are still available for allocation to medical subjects for populating with journals which currently appear in the Medicine (Miscellaneous) category.

There is considerable scope for improvement of the subject classification itself. For example, there is overlap and historic logic which requires re-examination at each level of the current ASJC table. For example,

- Environmental Sciences is classified as a Physical Science, while Agricultural and Biological Sciences are classed as Life Sciences.
- Many subjects in Health Sciences are classified in Life Sciences. For example, Psychiatry is listed as a subcategory of Health Sciences, while its partner subject, Psychology, is listed under Social Sciences.

## Discussion

The studies reported in this paper have been a valuable trigger to extensive collective reflection on the nature and purpose of the All Sciences Journal Classification scheme and its future development. At the most basic level, a change of name to the All (or Academic) Subjects Journal Classification scheme would more accurately reflect its incorporation of many non-STEAM journals in recent years.

Moreover, a broadening of the subject granularity to make use of the reserved and unused 4 digit identifiers would allow a more accurate allocate on many journals which are currently allocated to large repositories such as General Medicine (2700) and Medicine (Miscellaneous) (2701). This process could be combined more generally with a data cleaning exercise to reallocate journals which have seemingly been allocated the wrong code and to an inappropriate subject field.

Many authors have observed the limitations of current journal subject classification systems. Shir Aviv-Reuven and Ariel Rosenfeld (2024) (4) noted unusually sized categories, high overlap and incohesiveness between categories in both Web of Science and Scopus systems, and that across the two systems, journals are systematically classified to a different number of categories and most categories in either system are not adequately represented in the other system. They concluded that these irregularities and discrepancies were not anecdotal and could not be easily disregarded.

Mike Thelwall and Steven Pinfield (2024) (5) assessed the publication practices of specialist, cross-field and general academic journals against their Scopus classifications. They compared the Scopus subject fields of journals with the fields that best fit their articles' titles and abstracts, and also sought to distinguish between Scopus classification errors and misleading journal aims. They noted that some journals had titles and aims that do not match their contents, and that some topics were spread across many relevant fields. They concluded that such variations undermine citation-based indicators that rely on journal-level classification and may confuse authors in the search for appropriate journals in which to publish.

### **The Challenges of the “Multidisciplinary” Classification**

When all outlier journals have been reclassified and re-allocated to appropriate subcategories, there will still be a large number of journals which cannot be allocated to specific subject areas in consequence of their multidisciplinary content.

Multi-disciplinarity poses particular problems when attempts are made to define journals by their contents. This is highlighted in the ASJC by the necessary creation of “Multi-disciplinary” subsection in each of the four subject super-groups, to accommodate the many journals which are not otherwise classifiable. At the highest level, the four digit 1000 code has been allocated for journals which are so broad in their content coverage that they cannot be defined by any specific supergroup or subject classification.

The Multidisciplinary categories in the ASJC highlight the particular challenges of merging an academic subject classification with a journal classification. The vast majority of academic articles and texts focus on specific topics, through which they can reasonably be fitted into a logical subject classification. Therefore, “multi-disciplinarity” rarely applies or causes any classification problems at the article level.

### **The Levels of Multi-disciplinarity in Academic Journal Publication**

The “Multidisciplinary” label conceals a number of problems in subject classification.

It is not the purpose of this essay to explore such models in detail, but simply to point out that they require unique treatment in terms of classification of content. This classification problem arises where articles are bundled in the journal wrapper without any significant internal classification or segmentation.

A. At the highest level, the range of content may be seemingly so broad as to defy any form of subject classification, as with current “Mega-journals” such as the journals Heliyon, PLOSOne, or F1000Research.

B. At the next level of “Multidisciplinarity” are journals which provide wide subject field within any one super-group. High profile journals such as Nature, Science and Cell might be considered in such a class. In some cases, publishers of such journals have accommodated

the breadth of subject content in the flagship journals by moving to greater subject specialisation in the creation of topic-specific publications, such as “Nature Genetics”.

C. At the next level of “Multidisciplinarity” are journals which cover a wide variety of subjects within any one major subject area. Within my own subject field of Medicine are journals such as the British Medical Journal (BMJ), The Lancet; and The New England Journal of Medicine (NEJM), which provide with subject coverage within the Medicine classification , but with overlap into various other subject areas in the Health, Life and Social Sciences areas. Many university and institutional journals provide similarly broad content coverage on a smaller scale.

D. At the next level of “Multidisciplinarity” are journals which provide broad subject coverage within a particular professional or academic field of endeavour. For example, journals of Surgery or Medicine may cover all forms of Surgical or Medical Practice.

E. At the most subject-specific level, a journal will describe its intended content in its title and adhere to that intent in its actual content. Such journals are often affiliated to specialist societies, for example The European Journal of Heart Failure.

### **Multi-Disciplinary Content Mix and “Proportionality”**

A further problem with the classification of multi-disciplinarity lies in the quantitative mix of content from different subject fields, even where the cross-disciplinary mix of content is made explicit in the title, aims and scope of the journal. For example, the cross disciplinary journal Computational and Systems Oncology (E-ISSN 2689-9655) explicitly addresses both Computer Sciences (Physical Sciences) and Cancer (Medicine, Health Sciences).

In such a journal, multi-disciplinarity may vary at the article level between articles which address both subjects; to articles which address either subject field; so the article mix may vary considerably within that journal across the subject fields.

However we look at the problem of multi-disciplinarity, it is clear that it is very difficult to shoehorn many multi-disciplinary journals into a clear and simple subject classification.

### **Multi-disciplinarity and the Purpose of a Journal**

It is also helpful to understand the nature of multi-disciplinarity if we look at the implied and actual nature of the journal in which any article is published. This may be judged qualitatively on the basis of a range of factors which include its title, its implicit and its stated purpose, aims and scope, and the nature of the publisher. This leads into complex and challenging debate around the definitions of quality of both a journal and of a publisher, which are not the subject of this paper.

The purpose of a journal might be expected to be binding from its title and its stated purpose, aims and scope, there is also the challenge of distinguishing what may be defined as the Aspirational Intent of the journal in terms of subject coverage when it was created, and the Observed Content at the present time.

There may be evolutionary drift from the original purpose of a journal in consequence of its reputation for quality and the demand for space in its pages. Alternatively, there may also be deliberate drift from the original branding to increase manuscript flow for purposes of profit or prestige, whether through honest pursuit of quality or by malign intent.

Multidisciplinarity has been a long standing challenge in journal classification, In 1999, W. Glänzel (6) reported an item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. They noted a serious shortcoming of bibliometric studies in the *(Social) Science (s) Citation Index* through the lack of an universally applicable subject classification scheme. Subject classification of papers on the basis of assigning journals to subject categories failed for multidisciplinary journals such as *Nature*, *Science* and *PNAS*.

### **Article and Document Level Classification Systems**

For the reasons set out above, the complexities of merging a meaningful academic subject classification into a meaningful journal classification scheme have encouraged the exploration of article level classifications in SCOPUS, in the Web of Science and in other bibliometric reference systems. For example, the widely used Medline reference and article

search system has long used an article level classification rather than a journal level classification.

Singh et al (2020) (7) of the Department of Computer Science, at Banaras Hindu University, Varanasi, India, noted that the classification of research articles into different subject areas is an important task in bibliometric analysis and information retrieval, and that the recently introduced Dimensions academic database that uses article-based subject classification scheme to assigns the article to a subject category. They observed that of article-based subject classification did not prove superior to the journal-based subject classification, as used in SCOPUS and the Web of Science.

The evolution of machine learning (ML) and artificial intelligence (AI) systems provides the opportunity for the re-classification of journal content at the article based level, using the language of the titles and abstracts and the authorship any article or document. to discern the core theme of the article and hence to manipulate the information in various ways.

At a simple level, we have used the technique experimentally to re-allocate every article in the SCOPUS corpus to the nearest logical ASJC code. It was then possible to allocate the parent journal to the most appropriate ASJC code, based upon the analysis of the article collection in that journal. Clearly, this approach is constrained by the lack of subject granularity and the historic structure of the current ASJC.

A more sophisticated approach is to use the information generated by an article level analysis to propose a new version of the classification system, based upon and directed by such machine analysis and the subject range and diversity of content, whether sourced from Journals, conference proceedings or any other source. The parent publication of that content, whether a Journal, a book, conference proceedings, thesis or other source, will then be consistently classified by an updated version of the ASJC.

There is a rich literature on attempts to classify journal content and to compare proprietary classification systems using progressively more complex mathematical and statistical approaches at the level of article bibliometrics, as for example (Waltman and van Eck 2012,

Rivest et al 2021) (8, 9) . The latter noted equivalence in the performance of their deep learning approach with graph-based bibliometric approaches. However, they also noted that all machine learning approaches remained equivalent in their outputs to manual classification.

No consistent journal classification has yet emerged using Machine Learning, and all such approaches have significant limitations. These include the “common sense test”, whereby the trained human brain can process complex information and see solutions which may elude computer algorithms. We must also keep in mind that Bibliometrics are only a fractional measure of the content and of the utility of a journal. Other factors include the practical impacts of its content upon its readership; as for example in changing surgical practice or through other societal impacts.

Text analysis by machines is dependent upon the quality of the original writing and the clarity of communication of the titles, the abstracts and the full text content of an academic paper, may not necessarily be good exemplars of clear writing in their various native languages. Moreover, authors they may offer aspirational conclusions which do not accurately reflect the data presented in the papers which they summarise. Generative Artificial Intelligence summaries of the titles and abstracts may therefore not reflect the true data foundations and outcomes of a study.

The technologies of ML and generative AI are rapidly evolving, and it may well be that such a machine- directed methodology of classification will be created which is both persuasive and widely adopted. ML and AI readily address the challenges of scale, as we move from hundreds and thousands of journals to tens of millions of individual documents.

#### **Assignment of ASJC Codes based upon Author Profiles.**

A further modification of the postulated new approach to article level content classification is to use the author profiles which are associated with each article. The analytical model assumes that the authors of any article will be most closely and consistently associated with the subject classification, thus allowing the article to be allocated to the most appropriate

subject classification through author association. This hypothesis nevertheless requires rigorous testing, as many authors are multidisciplinary in their skills and interests.

### **Towards a Broader Characterisation of Academic Journals**

Despite its deficiencies, the ASJC will not be abandoned in the near future, not least as it represents a substantial financial and technical investment which is interlinked with other operational systems. We should nevertheless consider the properties of an updated version which would permit reclassification of the existing corpus of academic journals, and embrace the various and many other forms of formal academic outputs.

The content and purpose of an academic journal is defined by factors which include:

- The title, aims and scope of the journal
- The subject matter, whether consistent with the Title, Aims and Scope, or not;
- The authorship of the content of the journal;
- The professional inputs into the journal, including the editor(s) and the editorial board;
- The institutions, societies and associations which the journal serves, where appropriate
- The publisher of the Journal, and the characteristics of that publisher, including its ethics, its transparency, its governance arrangements and the jurisdiction in which its corporate and commercial operations are legislated
- The history and “geography” of the journal, in terms of its origins and the target communities from which it derives content and which it serves
- the metrics of the journal, including measures such as the number and types of article and content per annum, the citation activity, and its performance metrics vis a vis its immediate peer publications.

### **Quality Assurance of Documents and Journals**

Quality assurance in academic publication is a major challenge for document-based and journal-based classification systems. It has significance for the organisations, institutions and government agencies which depend upon them to help understand and compare their academic and research outputs at corporate, university, national and pan-national levels.

The factors and considerations which help us to evaluate quality fall outside the immediate scope of this paper, and will be addressed elsewhere.

## **The Importance of a Publisher-Based Component to any Journal Classification scheme**

The elephant in the room of academic publishing is the characterisation of publishers by their nature, motivations and behaviours. These in turn influence the policies and behaviours of the individual journals in their portfolios.

Jeffrey Beall, formerly a librarian at the University of Colorado in Denver until 2017, earned a lot of opprobrium through his efforts to characterise and publicise journal titles which caused him concerns. These arose from their quality, purpose and integrity in the matter of exploiting the “author-pays” open access model for profit without apparent regard to quality or coherence of the content. His term “predatory” has stuck, whether applied to individual journals or to their publishers, in the absence of a better description.

This is a complex field which does not fall within the remit of this paper. However, reference to it highlights the complexities of reflecting the ethical and trustworthy behaviour of the publisher in any journal classification scheme.

### **In Summary**

In this essay, I have considered the challenges of combining an academic subject classification with an academic journal classification system, and the current weaknesses to both approaches.

The “market leader” for such a system is the Elsevier All Science Journal Classification (ASJC) system. The ASJC has accrued various irregularities over some 25 years, which detract from its modern utility, overall credibility, fidelity and integrity. There is a strong case for major updating of the ASJC across all subject areas, and an accurate renaming as the All Subjects (or Sources) (and) Journal Classification scheme.

However, the core principles of ASJC system continue to be sound and useful, and adaptable to machine learning and artificial intelligence derived inputs. It offers the capacity for expansion and refinement of the subject codes and allocations without fundamental changes, major costs or operational disruption. There are plenty of “reserved” or unused

four figure codes which could be activated to increase the usability and granularity of the system, and modernisation can be introduced incrementally.

The significant weaknesses of the ASJC in its present form include the observations that:

- The legacy title no longer addresses the “All Subject” needs of the system.
  
- The subject coding system does not make full use of the available codes
- Within individual subject headings, many journals are seemingly misclassified.
  
- There is duplication among some subject categories. For example, ASJC Code 2709 (“Drug Guides”) serves little purpose and can be subsumed as category of Pharmacology.
  
- Duplication and mis-classification across top level subject fields is also evident. For example, many of the sub-classifications in Life Sciences include clinical elements which would be more logically sub-classified in Medicine, such as fields for Psychiatry in both Medicine (Code 2738) and Neurosciences (Code 2803).
  
- The heavy default use of a “Miscellaneous” option in each subject area is a clear indication of the lack of granularity in the current ASJC.
  
- Multidisciplinary journals of various types provide particular challenges to appropriate coding on a subject content basis.
  
- There are many user cases where it would be very helpful to have a much more granular search classification. For example, the subject area of Surgery (code 2746) provides one example of the lack of granularity in the ASJC. Surgery covers many specialities, including plastic surgery, vascular surgery, surgical oncology, colorectal surgery, and breast surgery, all of which subjects are covered by a number of specialist journals.
  
- the ASJC scheme needs to be more adaptable to emerging subject areas.

- Journals evolve in many ways, many of which are “legitimate” but some of which are “predatory”. The ASJC needs to be adaptable to significant changes in policy and content range in individual titles, which may need reclassification with journal evolutions.

The commercial considerations of corporate and organisational users of journal classification systems must be accommodated in any changes. The existing ASJC codes and classifications have become interwoven with a number of software systems and commercial products. The alteration of such systems and identification codes to reflect the evolution of journal policy and content may be impractical on cost and technical grounds.

### **Future developments with Source- and Journal based Classification systems**

At the document search level, researchers will invariably seek out articles by topic and purpose rather than by prior search for the journals in which they are published, as was once necessary in the era of paper. Journals themselves are now also challenged by ePrint and preprint servers and by other “direct to public” communication systems. However, for so long as journals, conference proceedings and books remain the principal vehicles for quality assurance in academic publication, there will be a need for a reliable classification system.

It therefore seems likely and necessary that one or other version of the ASJC will continue to evolve and adapt:

- with an emphasis on document level classification to optimise the allocation of journals to classification codes,
- with a better classification of journals which is based upon their observed behaviour rather than the aspirational statements of journals in terms of their publication practice.
- with the incorporation of the content of other publication modalities, including conference proceedings, books and book chapters, doctoral theses, patents, legal and policy documents

Jing Zhang and colleagues from the Chinese Academy of Sciences (2016) (10) have also sought to refine journal subject classification schema using measures of Journal coupling strength, which is a bibliographic measure of how closely related two documents are based on the number of shared references, and text mining of keywords.

Multidisciplinary journals will continue to challenge machine based classification, and Human expert assessment with a greater or lesser degree of machine teaming will remain important. It nevertheless seems likely that journal classification will move progressively from manual to machine assisted methods and to an automated ASJC classification assignment at the document level. This approach may also be used to create enhanced author profiles from the observed of patterns of publication and the bibliometrics of individual documents, rather than by virtue of the reputation and bibliometrics of the journals or other vehicles in which they are published.

This is an evolving discipline as new approaches and methodologies to the efficient and accurate classification of the global corpus of learning are explored and implemented.

### **Acknowledgements**

The opinions expressed in this essay are entirely those of the author and do not reflect or should be inferred as reflecting corporate opinion and policies of Elsevier BV. I am nevertheless grateful to Elsevier for inviting and supporting my participation in the SCOPUS journal evaluation programmes and discussions which have informed my thinking on this subject.

I am particularly grateful to Dr Rob Schrauwen of Elsevier for support and input into this essay, and for the data analysis and the figures.

## References

1. Web of Science Core Collection subject categories (accessed 28<sup>th</sup> October 2023)  
<https://webofscience.help.clarivate.com/en-us/Content/wos-core-collection/wos-core-collection.htm?Highlight=Subject%20Categories>
2. Wang Q and Waltman L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus,  
*Journal of Informetrics*, Vol 10, 2, 2016, pp 347-364, [doi.org/10.1016/j.joi.2016.02.003](https://doi.org/10.1016/j.joi.2016.02.003).
3. Stević A. Scopus is broken – just look at its literature category. blog post on Retraction Watch <https://retractionwatch.com/2024/07/17/scopus-is-broken-just-look-at-its-literature-category/>
4. Aviv-Reuven, S., Rosenfeld, A. A logical set theory approach to journal subject classification analysis: intra-system irregularities and inter-system discrepancies in Web of Science and Scopus. *Scientometrics* **128**, 157–175 (2023). <https://doi.org/10.1007/s11192-022-04576-3>
5. Thelwall, M. and Pinfield, S. (2024) *The accuracy of field classifications for journals in Scopus*. *Scientometrics*, 129 (2). pp. 1097-1117. ISSN 0138-9130
6. Glänzel, W., Schubert, A. & Czerwon, H.J. An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics* **44**, 427–439 (1999). <https://doi.org/10.1007/BF02458488>
7. Singh, P, Piryani, R, Singh, VK and Pinto, D.  
Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions  
*Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 2, pp. 2471-2476, 2020 DOI: 10.3233/JIFS-179906

8. Waltman L and van Eck NJ: “A new methodology for constructing a publication-level classification system of science” 2012 in Arxiv and Vol 63, The Journal of the American Society for Information Science and Technology (DOI 10.1002/asi.22748)

9. Rivest M, Vignola-Gagné E, Archambault É. Article-level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. PLoS One. 2021 May 2011; 16(5): e0251493. doi: 10.1371/journal.pone.0251493. PMID: 33974653; PMCID: PMC8112690.

10. Zhang J, Liu X & Wu L.

The study of subject-classification based on journal coupling and expert subject-classification system. *Scientometrics*, volume 107, pages 1149–1170 (2016)