Analysis of the ANZSRC for Use by the DOAJ

Martha Alvarez, Naomi Brown, Marion Thompson, & Rowan Russell

School of Information Studies (ÉSIS)

University of Ottawa

ISI 5302 – Knowledge Organization

Taught by Dr. Stefanie Haustein

December 11, 2022

Introduction

The Directory of Open Access Journals (DOAJ) is a prominent leader in the movement for open-access scholarly literature. It is an independent, non-profit organization made up of a small team of staff and over 100 volunteers from around the world (DOAJ, 2022). The DOAJ currently uses a truncated version of the Library of Congress Classification (LCC) system to organize its over eighteen thousand journals, but has identified several problems with this system. Namely, the LCC is too US-centric, outdated, and not granular enough to reflect its current needs. Dominic Mitchell (Operations Manager at the DOAJ) has identified specific key factors that are being sought in a new classification system to replace the LCC. The new system should be:

- Easy for both humans and machines to use
- Global to reflect the diversity of DOAJ
- Adaptable and extendable
- Easy to map and migrate from the existing system
- Open source

In this analysis we will review the Australian and New Zealand Standard Research Classification (ANZSRC) as a possible candidate for use by the DOAJ.

43	**HISTORY, HERITAGE AND ARCHAEOLOGY**		
	4301	Archaeology	
		430101	Archaeological science

Figure 1. Hierarchical structure of the ANZSRC

The ANZSRC actually contains three different classification schemes: Type of Activity (ToA), Fields of Research (FoR), and Socio-Economic Objectives (SEO). To reflect the purposes of the DOAJ as a directory of open-access scholarly research and the goal of replacing the LCC, we are limiting this analysis to the FoR codes. The FoR codes have a nested hierarchical structure comprised of Divisions, Groups, and Fields, with Fields being the most granular level. Figure 1 provides an example of this structure, where *43 History, Heritage and Archaeology* represents the Division, *4301 Archaeology* represents the Group, and *430101 Archaeological Science* represents the Field. In the newest version of the ANZSRC, there are 23 Divisions, 213 Groups, and 1967 Fields (CANZ, 2020). Figures 2 shows the Divisions, and Figures 3 and 4 illustrate the number of Group and Field codes in each Division.

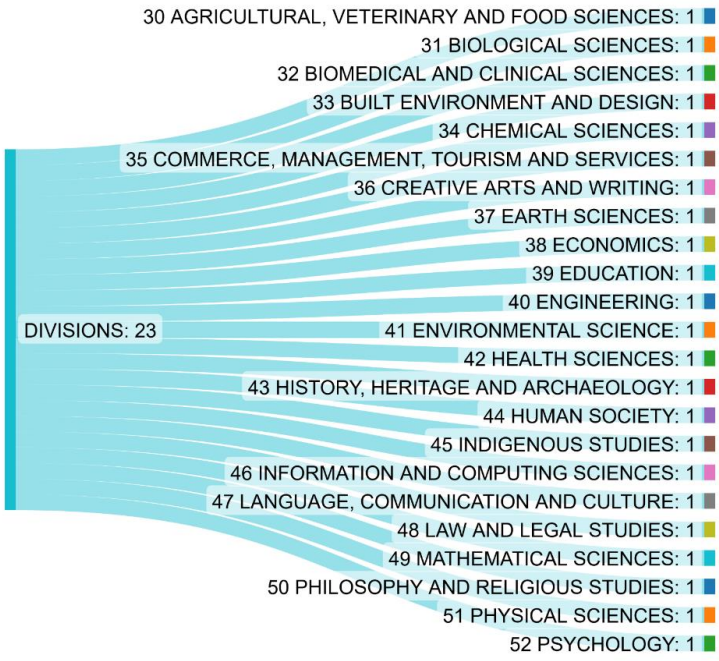

Figure 2. ANZSRC Divisions

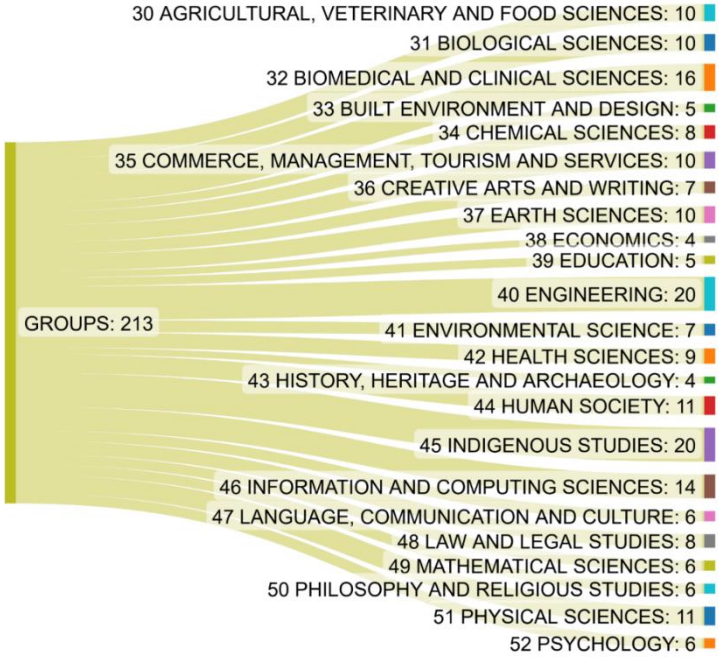

Figure 3. Number of ANZSRC Groups by Division

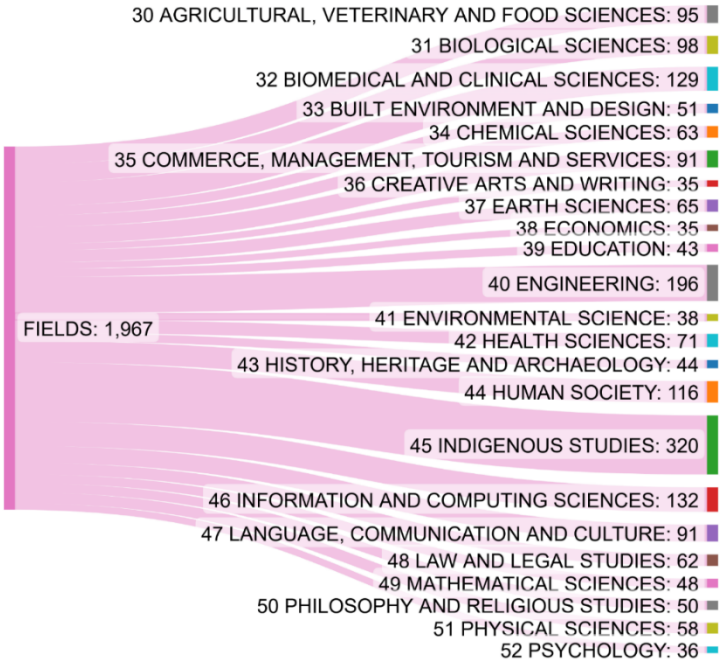

Figure 4. Number of ANZSRC Fields by Division

Literature Review

The ANZSRC

The following literature review provides a survey of relevant literature with respect to our assessment of the ANZSRC system as well as Dimensions, which we used as a primary reference source for how the ANZSRC could be adapted in another classification use-case. It should be noted that carrying out the literature review around the ANZSRC presented some challenges. Firstly, although it is a classification system that has its roots in the 1990s and was first officially implemented in 2008, there remains a relatively scarce array of academic publications with a focus on the ANZSRC. As a result, we relied heavily on non-academic sources such as evaluative documents written by the government entities involved, critiques of the system made by researchers and independent organizations, and a few others. Secondly, much of the literature available focused primarily on the implications of the ANZSRC's inclusion of indigenous categories of research. In fact, we found that all the resources we found were relevant to the ANZSRC either touched on or centered on the ANZSRC's work to include Indigenous research. While this is indeed important and will be touched on in our work, it is not the primary or sole focus for our purposes of evaluating the ANZSRC as a system appropriate for the DOAJ to replace their use of the LCC. However, it must be noted that this aspect is emblematic of ANZSRC and has implications that are valuable both to and beyond the Indigenous categories. Given the above points, the literature review will include both academic and grey literature sources.

For the purposes of this project, we drew on the Australian Bureau of Statistics (ABS) website page on ANZSRC as the source of our data for our analysis of the FoR portion of the classification system looking at the structure and content of the Divisions, Groups and Fields. It

also provided a starting point to familiarize ourselves with the basic outline of the ANZSRC. The source of the technical information on the latest release in 2020, the webpage includes the different classifications systems of the ANZSRC and provides open access to them in the form of various Excel format data cubes available for download. It also provides technical knowledge, concept definitions, notes some key points in the development of the system from its first release in 2008 followed by the 2019 assessment and subsequent 2020 release, and key stakeholders involved, but it does not provide conceptual analysis of the system. Important to note is the ABS (2020) statement that the ANZSRC is comparable to other international systems and that in their development they broadly adhered to the guidelines provided by Organization for Economic Co-operation and Development (OECD).

In an article which focuses specifically on the impact of the ANZSRC on the field of journalism, Nash (2020) also notes the system's capacity to be comparable on both national and international levels. Providing a history of the shifting placement of Journalism from the 1993 ASRC iteration up to ANZSRC's 2020 instalment, Nash (2020) offers an overview of the successes of, and challenges presented by, this shift as well as implications for the development of the field. He notes the decision to split Journalism from Communication and Media Studies and instead exist in the four-digit Group of Journalism and Professional Writing under the Division Language, Communication and Culture as ASRC was replaced by ANZSRC, and the subsequent need for members of the field to cultivate interdisciplinary relationships in order to establish funding opportunities and advance the field as a whole (Nash, 2020). Broadly, his analysis gives a clear example of the ways in which the ANZSRC both reflects and spurs academic development, and highlights interdisciplinarity as concept and practice.

In his comprehensive overview of the ongoing development of the ANZSRC, Hancock's (2022) view aligns with Nash's, pointing out that importance was placed by research communities in both New Zealand and Australia on ensuring categories would be created that would increase the visibility of and make space for indigenous research. He notes that the 2019 evaluation feedback indicated the existing categories established in 2008 were not sufficient and that the field of Indigenous research had changed since 2008, and remained unreflected in the ANZSRC (Hancock, 2022). This statement is relevant to both Indigenous research and a general evaluation of ANZSRC's level of relevance and need for review and updates (Hancock, 2022). Findings in the ANZSRC's Indigenous Research Consultation Draft (CANZ, 2019) confirmed that the 2008 version's lack of visibility for Indigenous research "negatively affected the ability of government, universities and other users…to report and analyze data" (CANZ, 2019, p. 12) and required addressing. Following consultations with over 500 stakeholders (Ngā Pae o te Māramatanga, 2021) including public consultations, a national workshop in Australia, targeted stakeholders including Māori and Pacific discipline experts in New Zealand, and many others, the 2020 release met these needs with the new FoR Division *45 Indigenous Studies*, and new SEO category of *Indigenous* (CANZ, 2020), and included the translation of all Māori codes into Te Reo, making them available in both Te Reo and English (Ngā Pae o te Māramatanga, 2021; CANZ, 2020).

In terms of spurring academic development, Nash argues that with the new Division *45 Indigenous Studies* and three new FoR codes of *Aboriginal and Torres Strait Islander, Māori and Pacific Peoples* for *Literature, Journalism and Professional Writing* "will precipitate major changes in funding, quality evaluation, research outcomes and coursework curriculum for both Indigenous and non-Indigenous Journalism" (2020, p. 139). The Ngā Pae o te Māramatanga

(2021) (New Zealand's Māori Centre of Research Excellence) at Auckland University also responded to this shift by acknowledging the disciplinary recognition and visibility, and state, "in a practical sense we will finally have Māori specific data with which to track investments, research impact and collective benefit" (Ngā Pae o te Māramatanga, 2021). Finally, Hancock (2022) seems equally satisfied that these concerns were satisfactorily met in the 2020 version, although he does say there remain other outstanding issues to be addressed in a future review.

These issues, which are interrelated, are how to address multidisciplinary and interdisciplinary research, the ANZSCR's adherence to mutual exclusivity, and the resulting lack of fluidity in representation of research. It has been acknowledged in the past that the ANZSRC's FoR codes are unable to "reflect the complex reality" (Bennett et al., 2011, p. 95) of this kind of research especially employed in the humanities and that this creates pressure to publish in certain journals (Bennett et al., 2011). The ANZSRC Outcomes Paper (CANZ, 2020) acknowledges that the consultation process revealed that a number of people were concerned about the way interdisciplinary and multidisciplinary research is represented in the system. However, upon further consideration of stakeholder feedback, the ANZSRC found that there is currently no other solution beyond assigning multiple codes to research data or dividing the research between multiple codes, and that this meets most needs (CANZ, 2020). And thus, this remains the system.

The argument made in the Outcomes Paper (CANZ, 2020) and further supported by Hancock (2022) rests on the 'classification best practice' of mutual exclusivity where different categories cannot overlap, must be in sequence, and according to the parent-child relationship; therefore, the creation of categories of *Multidisciplinary Research* and *Interdisciplinary Research* would step outside of those bounds (Hancock, 2022; CANZ, 2020). Hancock notes

these issues and suggests that future reviews of the system may eventually be able to address this challenge and provide more fluidity using “conceptual frameworks, or metadata modelling approaches within cloud-based IT systems” (Hancock, 2022). It should be noted however, that in 2009 an Australian research initiative, Excellence in Research for Australia (ERA), employed the ANZSRC system and chose to add a Multidisciplinary Code (MD) for their own purposes within their research initiative (Mamtora et al., 2014). They found that they assigned the category of MD to 3.4% of the ERA journals, and found it addressed some of the problems of classification (Mamtora et al., 2014, p. 268).

The response to the conclusion rested upon by the ANZSRC Outcomes Paper by some stakeholders was a polite push back to say that perhaps consensus was not reached, and as we know from the discussion above (Bennett et al., 2011; Mamtora et al., 2014), this is not the only concern raised on the subject with respect to the ANZSRC. In a piece written by the Network of Interdisciplinary and Transdisciplinary Research Organisations – Oceania (Nitro-Oceania), the authors argue that the ANZSRC draft conclusion did not “seem to reflect the majority of submissions made” (NITRO-Oceania, 2020, Feb, p. 1). To support their statement, they include an appendix of the public statements commenting on the topics of interdisciplinary, multidisciplinary or trans-disciplinary research and note that only 24% of respondents find the ANZSRC system “adequate” (NITRO-Oceania, 2020, Feb, p. 1). In short, they make several suggestions for this to be acknowledged in the final draft of the report (which do not appear to be included in the Outcomes Paper) and offer their assistance in meeting this complex challenge (NITRO-Oceania, 2020, Feb). Finally, they suggest that the core of the issue may be the system design itself, which is currently discipline-based, and urge ANZSRC to begin the process of

seeking out how to address this beyond the existing system (NITRO-Oceania, 2020, Feb), which echoes Hancock's suggestion.

While the ANZSRC does meet certain aspects of the qualification of "internationally comparable" (Nash, 2020) and was constructed with international standards documentation in mind (Hancock, 2022), we can see from these articles that the socio-cultural, historical, colonial, and geographic context of Oceania informs the organization and existence of its categories. While this is not an issue and is in fact valuable for its existing use by research communities in Australia and New Zealand who use the ANZSRC, including government, funding agencies, Crown Research Institutes, universities and others (Hancock, 2022), it is a bias. The DOAJ's existing classification system, the Library of Congress Classification (LCC), is equally steeped in the traditions, historical, social and geographical context against which it was conceived. In this case, the question is not whether a system without bias exists (it does not), but rather which biases are more problematic or can be tolerated, or as we indicate later, which systems can best be worked *with*. In brief, as the successful use (with some adjustments) of ANZSRC outside of Oceania by Dimensions and the Canadian Research Development Classification (CRDC) indicates, the system can be used outside of its original geographic home.

Although we note that although the DOAJ only has about 41 journals under keywords "indigenous" and 2 under "aboriginal" and this is but a very small fraction of the several thousand journals it houses, we suggest that this visibility and space-making processes, including extensive stakeholder consultations, for Indigenous research is crucial to developing the future of research and scholarship globally. Web of Science keyword searches show that Indigenous research has doubled in the last 10 years in Australia and New Zealand (CANZ, 2019, p. 7-8) and while the ANZSRC does not have categories for other Indigenous research, it has the

potential to provide the framework and inclusion for other Aboriginal and Indigenous methodologies and languages to be incorporated into it, perhaps more easily than other classification systems like the Dewey Decimal Categorization (DDC) or LCC.

As the Ngā Pae o te Māramatanga poignantly state, “Aotearoa New Zealand nurtures and works on unique areas of science, such as Indigenous plant species and soils that are not prevalent elsewhere and so the ANZSRC system helps the world understand what we are doing” (2021). This visibility and empowered information dissemination should be made available to other Indigenous researchers around the world, as well as other research categories, for that matter. Whether it is the category of Indigenous or other categories that are formed in ways that actively de-colonize, question Western and Eurocentric models of thought and information organization, this is a valuable part of what the ANZSRC brings to the table as a modern classification system.

Overall, the system presents a modern classification system option that provides creative and forward-thinking approaches with respect to the inclusion of Indigenous categories and subsequent visibility of research with broader positive outcomes; ability to classify research according to multiple fields; active approach to offering alternatives to Western and Euro-centric worldviews; stakeholder consultation; and is relatively up to date given its recent review process. Generally, the literature surrounding the ANZSRC is positive, although it highlights some challenges and difficulties with the existing system with the potential for some improvement. These include the potential for a bias based around the context of Oceania; the lack of consensus on how to deal with interdisciplinary and multidisciplinary research; and the unfortunate reverse effect of this on these fields contrasted with the positive visibility garnered for researchers in the Indigenous categories. The ANZSRC has already gone through one major review in its jump

from 2008 to 2020 and will likely go through another one in the years ahead. It remains to be seen how some of these issues, and potentially new innovations will come to the fore. As it currently exists, it could be an excellent classification system blueprint for the application to other research databases with some tweaks as the CRDC, Dimensions and ERA have done.

Dimensions

Dimensions is a research database which uses AI to classify articles and link them to relevant grants, clinical trials, and patents (Why Did We Build Dimensions, n.d.). The key element of interest for this paper is the fact that Dimensions classifies its articles using the ANZSRC system. It uses the FoR Division and Group levels, however it does not reach down to the most granular Fields level (Browse Fields of Research, n.d.). Using the Dimensions search function in the free version and filtering for "DOAJ" under the "Journal List," it was found that Dimensions shares over 12 million articles with the DOAJ. This is relevant because it illustrates the potential viability of using Dimensions metadata to help migrate DOAJ journals to the ANZSRC system. We also uncovered some potential issues as observed by others. For example, Zhang et al. (2022) discuss the problem of the Dimensions algorithm classifying some articles incorrectly. They note how an article about factors distorting scientists' decisions in the peer review process was categorized by Dimensions as *Applied Mathematics*, even though this is nowhere near the subjects tagged by the authors themselves. Zhang et al. continue on to suggest that this misclassification may be due to the presence of terms such as "models" in the abstract (p. 15). Bornmann (2018) had similar findings when looking at their own articles in Dimensions. They state that most of their papers seem to be misclassified, expressing a lack of faith in

Dimensions' accuracy (p. 639). These potential issues will be investigated further on in this paper.

Evaluation Summary

We used a three-part process to determine the compatibility of the ANZSRC system in classifying DOAJ journals. Two team members manually reclassified DOAJ journals and compared the results to the LCC classifications to determine subject accuracy. The other two team members used existing article metadata from Dimensions to reclassify the same set of journals in order to test the feasibility of using this metadata for large-scale migration. Additionally, we performed a one-to-one mapping of LCC to ANZSRC classes to further evaluate potential biases as well as to judge whether this could be a possible migration strategy. Results from these three approaches were analyzed to reveal the strengths and weaknesses of the ANZSRC system as a possible candidate for the DOAJ.

The random sample used in the first and second parts of our procedure were the same random sample of 100 journals from an Excel spreadsheet containing the metadata for over 12,000 DOAJ journals classified according to the LCC. In order to prevent cross-contamination of information from either classification process, each of these exercises was carried out by a different pair of team members.

Reclassifying DOAJ Journals Using ANZSRC

Procedure

For this section, we divided the 100 random journals equally between two group members, and each person proceeded with the manual re-classification according to the Divisions, Groups and Fields of the ANZSRC, limiting to about 3 codes or so due to time constraints for this paper. The goal was to focus on granularity.

In classifying the journals, we agreed that if a journal encompassed all the Fields within a Group, we would classify it as a Group, rather than classifying it under the *journal otherwise not classified* Field. We also agreed that should the journal be within the Group but not encompassed by the Fields available, then it would be classified under j*ournal otherwise not classified*.

Results

One of our findings was that the manual classification process was too time consuming. Two of the main reasons for this were connected to the wide scope of keywords associated with each journal in the metadata. First, the random sample was dispersed across a broad variety of fields with only some thematic overlaps, making the process of becoming familiar with a conceptual grouping, concepts, and new sets of terminology challenging. This was especially the case moving between vastly different fields. A set of similarly clustered journals, rather than disparate as was our case, would likely be quicker and easier to go through efficiently.

Second, was the discrepancy often noted between the keywords offered in the metadata and those found either in the description of the journal's website or, if not offered, in attempting to determine appropriate keywords based on an assessment of the articles present on the journal's

website. A more consistent set of keywords would be helpful not only in this case, but perhaps also for researchers looking either for relevant research or for the most relevant journal in which to publish according to key terms. We might suggest a dual system of pre-set keywords complemented by open tagging by researchers and authors. This would hopefully cover a set of more formal and limited keyword terminology for authors/journal editors to choose from when submitting their journal to the DOAJ, while also allowing a broader system of tags that could provide flexibility and fluidity in fields as terminologies change or where journals have a wider scope of themes.

Beyond the mechanics of the process itself, we also found several interesting insights on the ANZSRC through the manual classification and comparison with the truncated LCC used by the DOAJ. First, we discovered that there is no “war” or “military” section as is found in the DOAJ, but there was *Defence Studies*. We hypothesized that this was an indication of the choice to move away from a Western, Euro-centric model of looking at the world that positions war as aggressive rather than defensive and uses language to signify this. Further, while New Zealand and Australia do have a history of involvement in several military conflicts (for example World War One and Two; national conflicts), these may be considered relatively recent by comparison to the long military histories of Europe, the Middle East and Africa, and so the question of socio-geographic relevance may also be a factor. The existence of *440804 Defence Studies* did provide a category as necessary.

Several discoveries highlighted the presence of a context couched in what is relevant to Oceania more than what would be necessary to meet globally appropriate research data classification. For instance, there was no category for Latin American Cultural Studies or Literature, and no satisfactory alternate category, leaving a gap. We found that there were a lot of

different categories focusing on agriculture, agronomy and food science; a clear reflection of the agrarian-based economies of Australia and New Zealand. In fact, these Groups and Fields seemed to overlap, making it difficult to place journals for classification. We suspected that this might reflect field-specific knowledge, where an individual working in these domains would understand and benefit from these granular divisions, as well as be familiar with and guided by the keyword terminology. A similar trend was noted in the Division of *Engineering* which quickly becomes extremely specific at the Field level.

Specific terminology within Groups were at times displaying this granularity and again, without necessarily field-specific knowledge, slowed down the process to categorize certain journals as we looked up terminology to ensure accuracy. For instance, *Earth Sciences* and *Geosciences* were category terminologies that had to be looked up and compared in order to understand best journal placement. It's also likely that the ANZSRC better reflects the evolution and development of different fields and more recent ways of talking about these fields whereas the older version of the LCC used by the DOAJ remains pinned to older vocabularies and by extent, ways of thinking of the world. While some struggles may exist in the process of an initial shift from LCC to ANZSRC, it is likely that the granularity and updated catalog the latter provides may be well worth it with respect to the representation of some fields of research.

By contrast to this specificity, we felt the absence of categories for *General Science* and *Technology*, the latter of which was removed for the 2020 version after ANZSRC's assessment and consultation with stakeholders (CANZ, 2020). These differences in technology and field were especially noted given the DOAJ's categories of *Technology* and *Engineering* as much broader categories, making it easier to lump multiple concepts together.

A distinct advantage of the manual classification was the granularity, and the ability to assign multiple codes to each journal, which was often necessary to properly reflect the content of the journals, and ensure they are findable. We found that out of the 100 journals, 50% were assigned three or more codes, while journals classified using either one or two codes made up about 25% each. Given that the ANZSRC Outcomes Paper (CANZ, 2020) details that their solution to meeting the needs of multidisciplinary and interdisciplinary research is to assign multiple codes, the migration of the DOAJ journals to ANZSRC would need to be according to a system that could ensure the assignment of multiple codes. If our sample of journals is at all indicative of a likely trend among DOAJ journals, our finding that about 75% of journals were best represented using two or more codes, affirms this importance.

In sum, despite the advantages of depth, insight, and accuracy afforded by the granularity achieved through manual classification, the length of time required to complete 100 journals does not indicate a feasible undertaking for manual classification for the entirety of the DOAJ database migration. We therefore do not recommend this method as a stand-alone approach.

Limitations and Discussion

The manual system offers some advantages, especially in terms of insights and the capacity to assign multiple codes to journals, thereby representing the research content far more accurately, and rendering it more accessible. However, as we noted, it is too time-consuming to be feasible. The ideal solution would be to build and train an AI program that could carry out this depth of task without requiring the same volume of hours of human labour. Alternatively, in

order to suit the DOAJ better, we could consider a different classification system that is more similar to the existing LCC, making the shift more easily one-to-one.

If we were to remain with a human-based approach, the application of more consistent descriptor keywords to journals at the time of submission would speed up this process. Similarly, given the DOAJ is made up of volunteers, manual classification participants could be organized into groups according to their own academic/research experience and be assigned to journal fields more relevant to their existing knowledge. This would likely increase the speed at which classification could take place.

Based on our data sample, we found that the ANZSRC would accurately classify a significant portion of the journals in the DOAJ. Noted discrepancies would likely be the categories of Indigenous research failing to reflect non-Oceania Indigenous scholarship, and at least indicated by the relatively small sample, research areas not directly relevant to the Australia/New Zealand context, as we found in the case of a lack of category for contemporary Latin American Culture and Literature. This reflects the findings of the literature review, wherein bias exists toward the context of Oceania. Again, while bias is inevitable, the question remains whether this bias can be worked with by the DOAJ and be worthwhile to address by supplementing the ANZSRC with additional Divisions, Groups and Fields to meet a more international use-case, or not. While many journals would be classified, the current scope of journals in the DOAJ would not be fully classified in the ANZSRC without some additional Groups and Fields, and maybe even Divisions.

Despite these challenges, the ANZSRC provides a refreshing approach from the LCC, couched in much older and arguably more problematic biases. With its focus on Indigenous research in Oceania, it looks to the future of visibility and recognition for other Indigenous

research around the world. The increased granularity in many other Groups and Fields also provides distinction for many areas of research to be recognized. As evidenced by the focus on Indigenous research, but also seen in the case of other fields, for instance categories focused around war being replaced by *Defence Studies*, the ANZSRC provides a much-needed destabilization of the Western, Euro-centric model of conceiving of the world.

Dimensions-Assisted Reclassification

If the DOAJ wishes to proceed with replacing the LCC with the ANZSRC, a process for moving the journals over to the new system would have to be conceived. Given Dimensions' existing use of ANZSRC, we considered whether its metadata could be used to help with the mapping process. In seeking a solution, we sought to determine whether the approach would be accurate and comprehensive enough to mass migrate DOAJ to ANZSRC. We compared our manual classifications with those found in Dimensions to determine whether the Dimensions classifications would be accurate and therefore useful. Our rationale for following these two main classification exercises was to compare manual classification as a more human-based, traditional cataloguing method with a faster, AI driven approach.

As mentioned, Dimensions is a research database that shares over 12 million articles with the DOAJ, and uses the ANZSRC's Fields of Research (FoR) classification system to classify its articles. This strategy would make migrating the DOAJ's journals from the LCC system to the ANZSRC system much easier and less labor-intensive, as the DOAJ staff and volunteers could use the existing metadata from Dimensions instead of re-classifying thousands of journals manually.

Procedure

We performed the assisted classification procedure using the same random sample of 100 journals as was used in the previous method. We divided the 100 journals between two group members, so each member re-classified 50 journals. We searched for the journal titles one-by-one in Dimensions under “Source Title” and then limited the results in order to display the articles contained in each journal. We also limited the results to the date range 2019-2021. On the right side of the page, under “Research Categories”, the ANZSRC code assigned to the highest number of articles is displayed at the top. We then classified the journal under this code, under the deliberate assumption that if most of the articles in a journal are classified with this code, it would be an appropriate classification for the journal as a whole.

Overall, the codes for our 100 sampled journals were mostly a match when comparing the Dimensions-assisted reclassification method and manual classification method. 78% of the journals were a match, while 16% were a partial match, and 9% were not a match. The remaining

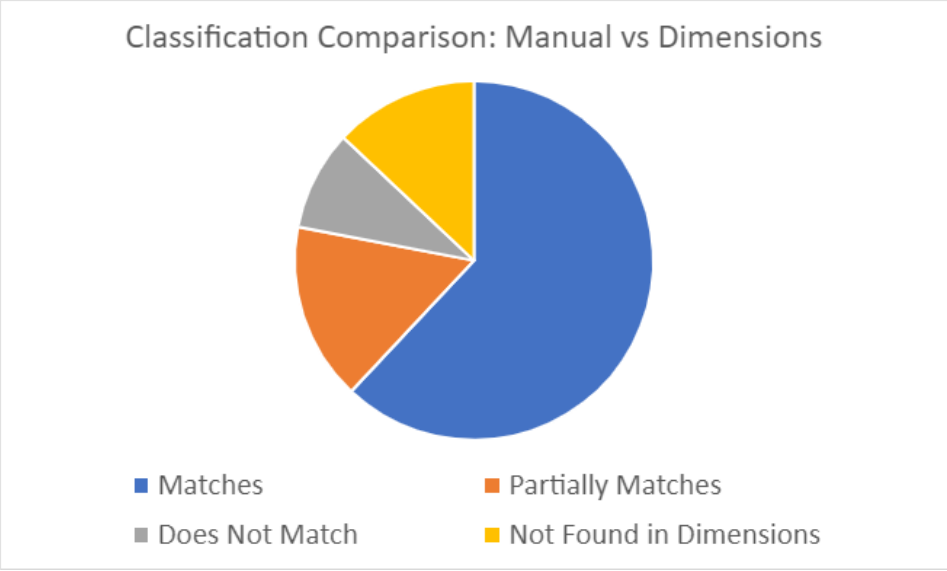

Figure 5. Comparison of manual and Dimensions-assisted reclassifications

13% of the journals were those that could not be found in Dimensions and therefore only had a manual classification.

We defined a “match” result as when the Dimensions code was within the same Division as all manual codes assigned, and is thus both accurate and comprehensive. A “partial match” was when the Dimensions code was within the same Division as some of the manual codes, but not all of them, meaning the Dimensions code is mostly accurate but not comprehensively so. When the Dimensions code did not match any of the manual codes, we indicated “does not match” and of course sometimes the journal simply could not be found in Dimensions at all. Figures 6 and 7 help to illustrate the distinctions between matches and partial matches with examples.

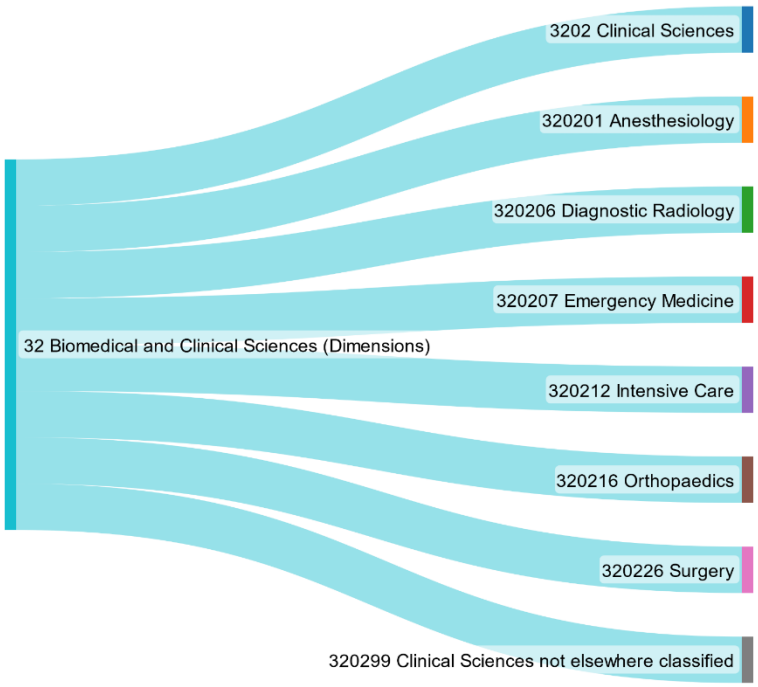

Figure 6. Full Matching for *Trauma Case Reports* journal

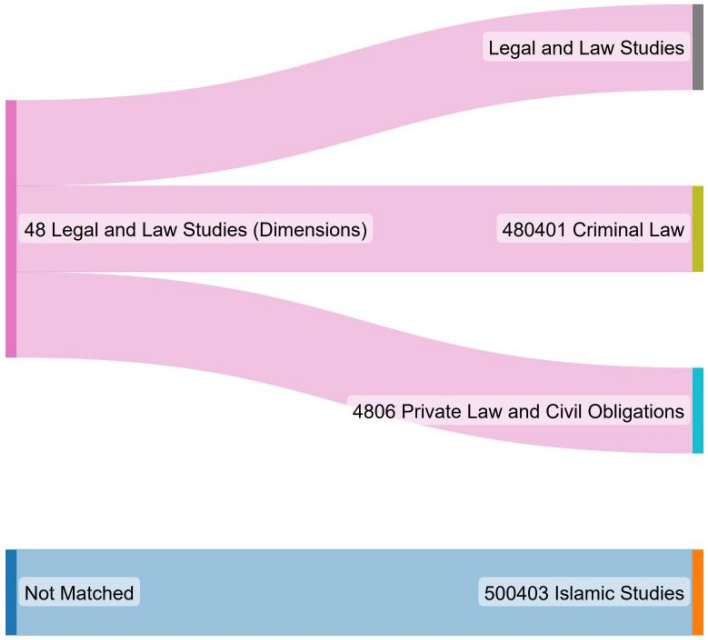

Figure 7. Partial Matching for *Yustisia* journal

Of the 9% of our sample journals that did not match in classification, it was the Dimensions classification that was incorrect in every case. However, this percentage is fairly small overall, and could be solved with a human screening process. Finally, almost all of the Dimensions-classified journals were classified with a two-digit ANZSRC Division, while the manually-classified journals had much more granular classifications, down to Groups and Fields. This shows that using the Dimensions method could lead to classifications that are too broad and do not make use of the well-detailed categories that are an asset of the ANZSRC system.

Based on our investigation and comparison of the two approaches, we determine that yes, using Dimensions metadata is accurate enough to offer the possibility of mass migration of the DOAJ to ANZSRC as it would be significantly faster than doing a complete from scratch migration. However, the results indicate that this would have to be carried out in combination

with manual human involvement as Dimension's results are not granular enough or consistently accurate. As a result, additional time and labour resources would need to be allocated for this approach, and partial manual classification would still be necessary.

Limitations

Along with the 13% of our journals that could not be found in Dimensions, we encountered some further problems during the assisted re-classification process. First, newer journals often had articles that were not yet classified, and sometimes articles were not classified in general. Second, especially in the case of multidisciplinary journals, using the most general classification could mean using an inaccurate classification for that journal. Third, and finally, the main research category for a journal on Dimensions was sometimes incorrect with no clear reason why, such as in the case of one of our journals, *Jurnal Keperawatan Indonesia*, a journal about nursing, which is classified on Dimensions as ANZSRC code *50 Philosophy and Religious Studies*.

LCC to ANZSRC Direct Remapping

We also decided to do a remapping of the ANZSRC classification against the LCC system that is currently used by the DOAJ. The purpose of this was to see if it was possible to do a one-to-one mapping or migration of subject headings, and to more accurately see any of the biases or limitations of the ANZSRC by comparing the FoR section directly to the LCC Classes and Subclasses.

Procedure

We used the LCC Classes and Subclasses as the basis of our mapping and mapped the ANZSRC FoRs against it, trying to stay at the Division or Group level, which are the same equivalent levels of classification. We then went through each class and subclass and using the FoR metadata provided by the ANZSRC, utilized the Ctrl + F function to locate similarly named ANZSRC Divisions or Groups. In some cases, we had to go down to the Field level, as there was no higher class that was an appropriate fit. Where no appropriate ANZSRC Division, Group, or Field could be found, "N/A" was placed in the table. Mapping was done based solely on the names of the LCC Classes and Subclasses, and ANZSRC provided metadata, without any outside influence from the DOAJ website, journal keywords, or other literature sources.

Results

While some subject headings were able to be adequately re-mapped directly to the ANZSRC, there were other parts of the LCC where certain Subclasses did not have their own direct match and were joined together by other Subclasses into the same ANZSRC Division, Group, or Field. There were also instances in which no equivalent ANZSRC classification was found at all. For example, with the *R Medicine* section of the LCC, most of the Subclasses could be remapped to the Groups or Fields of the ANZSRC that fell under the *32 Biomedical and Clinical Sciences Division*. This was an instance in which the one-to-one mapping worked well across both classification systems.

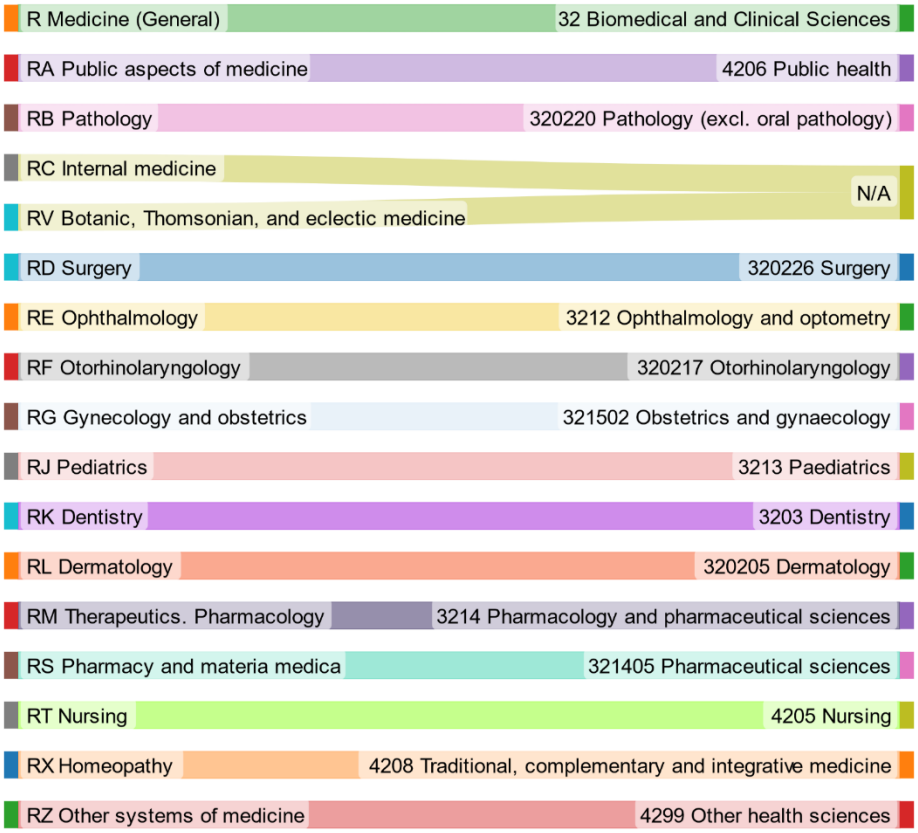

Figure 8. *R Medicine* Mapping Results (high success)

However, in the *D History* LCC Class, each individual European country history Subclass was grouped under the ANZSRC Field *430308 European History (excl. British, classical Greek and Roman)*, as the ANZSRC does not possess individual categories or each European country or even region. The LCC, on the other hand, did not have as many options for Oceanic and Indigenous history.

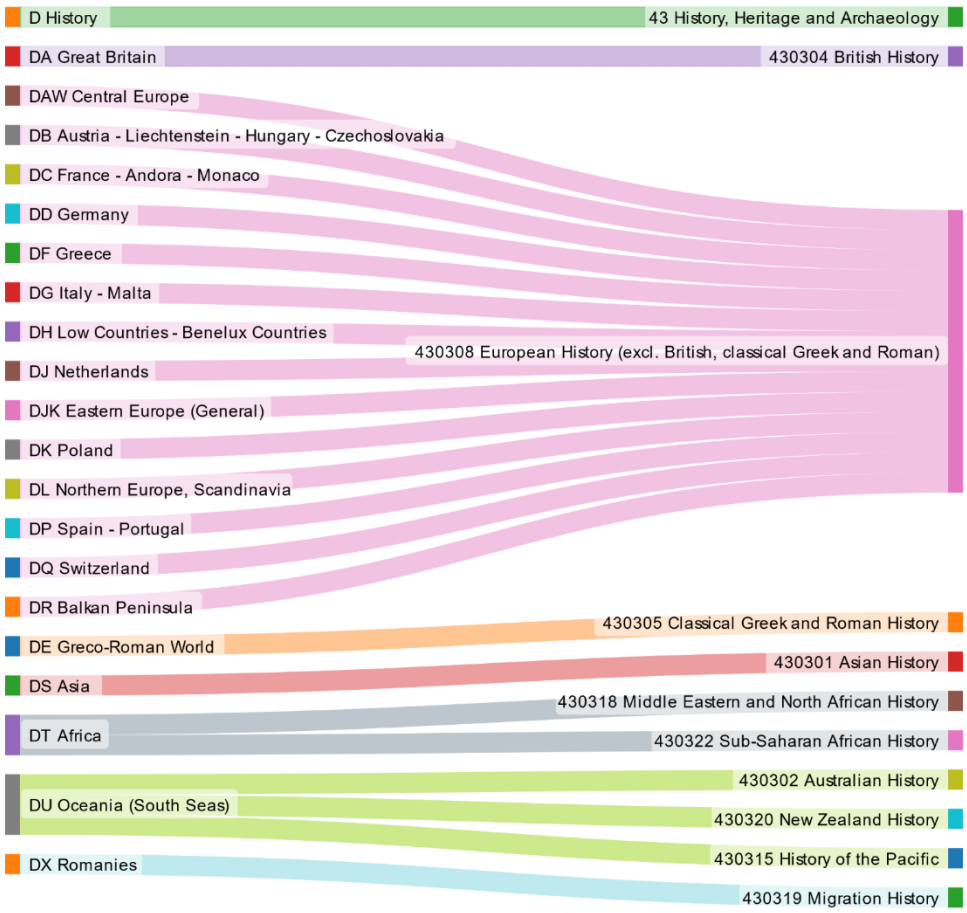

Figure 9. *D History* Mapping Results (partial success)

As for the *U Military Science* and *V Naval Science* sections of the LCC, there were almost no ANZSRC equivalents whatsoever. There is no equivalent ANZSRC Division for *U Military Science*, and the closest fit for this LCC class was the ANZSRC Field *440804 Defence Studies.* The only other option for the other Military Sciences LCC Subclasses would have been for them to be grouped in under the *Defence Studies* Field as well. There were, however, ANZSRC Field equivalents for Navigation and Naval Architecture, which made sense that both Australia and New Zealand are island nations.

Figure 10. *U Military* and *V Naval Science* Mapping Results (low success)

It also became apparent during the mapping process that we would be unable to stick to the Division and Group level for our mapping of the LCC Classes and Subclasses, as there were often no appropriate one-to-one matches for these in the ANZSRC. In many cases, it was necessary to use categories in the Fields level in order to find a better fit. This, along with the fact that the ANZSRC system is also far more granular than the LCC and uses a different system of organization with an emphasis on more Australian and New Zealand perspectives, would make a perfect one-to-one migration nearly impossible to do. It is also important to note that the

LCC also has no Class or Subclass for Indigenous studies, culture, or research, and still has a strong focus on Western European values. It also uses outdated terms like Oriental, and lumps together non-European cultures and languages under the same Subclass.

Discussion and Limitations

Each classification possesses biases and is shaped by the culture and geographic location it originated in. The LCC is shaped by European and American (as well as white, Christian, and colonialist) values that have shaped European and American society over the last hundred or so years, while the ANZSRC is a newer system influenced by Australian and New Zealand geography and politics, that is also actively attempting to dismantle colonialist biases by providing more focus on Indigenous subjects. This has influenced the organization and hierarchy of the ANZSRC Division, Group, and Field subjects. It appears as though some ANZSRC categories would be used more heavily than others, and that there are still whole sections that would be lacking for certain LCC subjects such as European History and Military studies. Furthermore, the ANZSRC system separates out into far more Groups and Fields than the LCC has Subclasses and Bottom Level Classes, making them structurally dissimilar from one another, and creating further incompatibility for one-to-one migration of Classes from the LCC to the ANZSRC.

Recommendations and Adoption

The ANZSRC system is a modern classification system that has many qualities that make it stand out in comparison to the other classification systems in use around the world today, such

as its granularity through detailed research categories, its ease of use with clear and simple numerical classification codes that are easy for both humans and computers to interpret, and it still has lots of room to grow by means of adding more categories. It is also an inclusive system that made a big effort towards highlighting Indigenous research by giving Indigenous works their own Division, which hopefully sets an example for other classification systems to follow. It also separates British history from the rest of European history in its research categories in an anti-colonialist effort to destabilize the longstanding trend of British history being dominant over the histories of other European nations in the academic world. The ANZSRC also offers a system with a non-Western or Euro-centric worldview that is crucial for reducing bias in the world of academia. However, it is not the best system for international use as it is, of course, mainly focused on the areas of Australia and New Zealand. Adding numerous new categories could remedy this in the future, and while this system has the space to grow and accommodate these new categories, it would be a grand undertaking. Considering the pros and cons of the ANZSRC classification system, we do not recommend it for adoption by the DOAJ at this time, but rather we suggest that the DOAJ take into consideration some of the benefits of the ANZSRC in its search for a new system. We also pose the idea that a modified version of the ANZSRC system could be adopted by the DOAJ with the addition of new categories custom-tailored to suit their international focus, perhaps with the help of AI or machine-learning.

There were several challenges we came across in the process of testing the use of the ANZSRC system for the DOAJ that led to our conclusion to not recommend the ANZSRC for adoption. First, the ANZSRC system still struggles to accommodate multidisciplinary works. These works will have to be categorized under multiple divisions in order to be easily accessible by researchers. Second, we found that some areas of research are still not represented in their

own research categories. For example, there is no category for Medieval history. This could be because the Medieval era was not a significant time period for Australia and New Zealand, so there was more of a focus on highlighting Indigenous histories, like that of the Māori peoples, in the classification system instead. However, there still needs to be somewhere to classify Medieval works, such as one of our sample journals *The Journal of the Spanish Society for Medieval English Language and Literature*. There were also no categories for military subjects, or certain non-Oceanic cultural subjects like Latin American culture. These had to be classified in the language or history categories, which are not the most appropriate placements. Third, the DOAJ's current system has some issues that posed challenges when attempting to classify its journals with another classification system that will need to be improved. For example, authors often assign keywords to their own works using a variety of different wordings that make it difficult to achieve consistency across the system (e.g. earth science, geoscience). This means that extra time will have to be spent during the classification process to look up keywords to find matches. We suggest a set list of keywords that could be selected from during the journal submission process to increase consistency and efficiency in the classification process. We also suggest an improved visibility of the classification system in the DOAJ so that authors will be familiar with what research categories exist, helping guide the decisions they make when classifying their own works. As a potential future consideration, machine-learning, such as a smart algorithm or an AI, could also help with this issue by reading the documents and placing them in the correct classification, similar to Dimensions. A final option could be for DOAJ volunteers to be coordinated strategically to classify works in research areas that they are familiar with to improve efficiency and save time.

With all these points taken into consideration, we do not recommend the adoption of the ANZSRC system for the DOAJ at this time. However, with its ample room available for new categories to be added, it could be a good system for the DOAJ in the future if the work needed to be done to add new categories is feasible. It is also important to note that while the ANZSRC has yet to incorporate several non-Oceanian cultures into its research categories, its efforts to disentangle European and Oceanic academic research from the British imperialist lens through which they have been viewed for a long time, especially with its inclusion of a transnational history and Indigenous categories, makes it a classification system that has great value. It sets a good example that the DOAJ and other classification systems of the world could follow.

Conclusion

In summary, we conducted a three-part evaluation to determine the compatibility of the ANZSRC in classifying DOAJ journals which are currently classified using the LCC system. First, manual reclassification of DOAJ journals was performed using the ANZSRC to examine the scope and characteristics of the system. Then, we investigated the accuracy of generalized Dimensions classifications by comparing it with our manual codes to confirm if this existing metadata could be used by DOAJ to aid in migration, should the ANZSRC be suitable. Lastly, we mapped ANZSRC codes against the LCC system to further evaluate the suitability of the systems and test if one-to-one mapping was a viable migration solution.

Judging from our results, we unfortunately do not recommend the ANZSRC as it currently stands for use by the DOAJ, due to its difficulties in properly representing multidisciplinary research, as well as the fact that it reflects a particularly Oceania-specific view. That said, the ANZSRC provides a strong example of a modern classification system for

research, displaying excellent granularity and a refreshing departure from the US-centric imperialist view of the LCC. The DOAJ is open to consider using a modified version of the ANZSRC that would more adequately suit its global scope.

References

Australian Bureau of Statistics. (2020). *Australian and New Zealand Standard Research Classification (ANZSRC)*. https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc/latest-release

Australian Bureau of Statistics. (n.d.). *Australian and New Zealand Standard Research Classification (ANZSRC)*. https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc

Bennett, D., Genoni, P., & Haddow, G. (2011). For codes pendulum: publishing choices within Australian research assessment. *The Australian Universities' Review*, *53*(2), 88-98. https://doi.org/10.3316/ielapa.388269150502258

Bornmann, L. (2018). Field classification of publications in Dimensions: A first case study testing its reliability and validity. *Scientometrics*, *117*(1), 637–640. https://doi-org.proxy.bib.uottawa.ca/10.1007/s11192-018-2855-y

Chambers, T. (2019). Dimensions. *The Journal of the Canadian Health Libraries Association = Journal De l'Association Des Bibliothèques De La Santé Du Canada*, *40*(1) https://doi.org/10.29173/jchla29405

Commonwealth of Australia and New Zealand (CANZ). (2019). *Indigenous Research Consultation Draft: Australian and New Zealand Standard Research Classification Review 2019.* Australian Bureau of Statistics, Stats NZ, Australian Government, & Ministry of Business, Innovation and Employment. https://www.mbie.govt.nz/dmsdocument/7473-consultation-draft-indigenous-research-australian-and-new-zealand-standard-research-classification-review-2019-pdf

Commonwealth of Australia and New Zealand (CANZ). (2020). *Outcomes Paper: Australian and New Zealand Standard Research Classification Review 2019.* The Australian Research Council, Stats NZ, Australian Government, & Ministry of Business, Innovation and Employment. https://www.arc.gov.au/sites/default/files/anzsrc_review_outcomes_paper_v1.1.pdf

Dimensions. (n.d.). *Fields of Research (ANZSRC 2020)*. https://app.dimensions.ai/browse/categories/publication/for?and_facet_for=80045

Dimensions. (n.d.). *Why did we build Dimensions?* https://www.dimensions.ai/why-dimensions/

Directory of Open Access Journals (DOAJ). (2022). *Find open access journals & articles.* https://doaj.org/

Hancock, A. (2019). *Australian and New Zealand Standard Research Classification (ANZSRC).* International Society for Knowledge Organization. https://www.isko.org/cyclo/anzsrc

Hazlett, H. (2021, April 27). *Updates on Recognizing and Supporting Indigenous Research from New Zealand and Australia.* DORA. https://sfdora.org/2021/04/27/updates-on-recognizing-and-supporting-indigenous-research-from-new-zealand-and-australia/

Legendre, A. (2019). *Canadian Research and Development Classification* (CRDC). International Society for Knowledge Organization. https://www.isko.org/cyclo/crdc#5.2

Mamtora, J., Wolstenholme, J. K., & Haddow, G. (2014). Environmental sciences research in northern Australia, 2000–2011: A bibliometric analysis within the context of a national research assessment exercise. *Scientometrics, 98*(1), 265–281. https://doi.org/10.1007/s11192-013-1037-1

Ministry of Business, Innovation & Employment. (2020). *Review of the Australian and New Zealand Standard Research Classification.* https://www.mbie.govt.nz/have-your-say/review-of-the-australian-and-new-zealand-standard-research-classification/

Nash, C. (2020). Gentle sounds, distant roar: A watershed year for journalism as research. *Pacific Journalism Review: PJR, 26*(2), 132–141. https://doi.org/10.3316/informit.569088226794396

Network of Interdisciplinary and Transdisciplinary Research Organisations (NITRO) – Oceania. (2020, Feb). *Response to Consultation Draft of the Australian and New Zealand Standard Research Classification (ANZSRC) Review 2019*. https://nitro-oceania.net/wp-content/uploads/2020/02/NITRO-Oceania-2020-response_ANZSRC-review.pdf

Ngā Pae o te Māramatanga. (2021). *Acknowledging Significance of Indigenous Scholarship*. https://www.maramatanga.co.nz/news-events/news/acknowledging-significance-indigenous-scholarship

Zhang, L., Sun, B., Shu, F., & Huang, Y. (2022). Comparing paper level classifications across different methods and systems: An investigation of Nature publications. *Scientometrics*. https://doi.org/10.1007/s11192-022-04352-3