LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 32, (2022) 1–22 | e-ISSN: 2213-056X

The Curtin Open Knowledge Initiative: Sharing Data on Scholarly Research Performance

Katie Wilson

Te Herenga Waka Victoria University of Wellington, Wellington, Aotearoa, New Zealand
katie.wilson@vuw.ac.nz, https://orcid.org/0000-0001-8705-1027

Lucy Montgomery

Curtin University, Perth, Australia
lucy.montgomery@curtin.edu.au, https://orcid.org/0000-0001-6551-8140

Cameron Neylon

Curtin University, Perth, Australia
cameron.neylon@curtin.edu.au, https://orcid.org/0000-0002-0068-716X

Chun-Kai (Karl) Huang

Curtin University, Perth, Australia
karl.huang@curtin.edu.au, https://orcid.org/0000-0002-9656-5932

Rebecca N. Handcock

Curtin University, Perth, Australia
rebecca.handcock@curtin.edu.au, https://orcid.org/0000-0001-5903-6620

Abstract

In the current era of worldwide competition in higher education, universities are caught up in market processes that encourage compliance with the measurement systems applied by world university rankings. Despite questions

Uopen Journals | http://liberquarterly.eu/ | DOI: 10.53377/lq.11544

about the rankings' methodologies and data sources, universities continue to adopt assessment and evaluation practices that require academic researchers to publish in sources indexed by the major commercial bibliographic databases used by world rankings. Building on a critique of the limited bibliometric measures and underlying assumptions of rankings, the Curtin Open Knowledge Initiative interdisciplinary research project aggregates and analyses scholarly research data including open access output from multiple open sources for more than 20,000 institutions worldwide. To understand who is creating knowledge and how diversity is enacted through the transmission of knowledge we analyse workforce demographic data. In this article, we discuss the project's rationale, methodologies and examples of data analysis that can enable universities to make independent assessments, ask questions about rankings, and contribute to open knowledge-making and sharing. Expanding on our presentation to the LIBER Online 2021 Conference, we discuss collaboration with academic libraries and other scholarly communication stakeholders to develop and extend the open knowledge project.

Keywords: open knowledge; open access; research performance; diversity; gender

1. Introduction

Critiques of world university rankings and their processes, indicators and proxies, including prejudices and blind-spots in relation to the measurement tools and data sources have grown since 2007. For example, the databases providing publication and citation data consist of predominantly English language, elite publication sources whose scope precludes research output from many countries and marginal communities. As universities are driven by the rankings to compete, their focus becomes less on national and local needs, a major shift (Hazelkorn, 2018), reducing publication in local sources that are accessible and relevant to their communities. Rankings persist, creating their own reality and paradoxically changing university strategies and performance (Gadd, 2021).

The Curtin Open Knowledge Initiative is a strategic research project that aims to move beyond university rankings, to promote openness in institutions and research production (Montgomery et al., 2021). The project aggregates and analyses multiple dimensions of research output that can facilitate dialogue and encourage different ways of thinking about research impact. Central to

the project's approach is gathering together publicly available data that is often fragmented. This includes bibliographic, affiliation, citation, funder and workforce demographic data from a range of public sources to understand research performance including open access, collaborations, citation advantages, publishing, funding and institutional diversity.

The paper falls into four parts. The first discusses the rationale and background to the project including a critique of world university rankings and the genesis and development of our dataset and dashboard tools. In the second section we discuss methods used in developing the project's dataset, data sources and analysis, and a feedback survey undertaken with academic libraries. The third section discusses two key aspects of the project: open access research analysis and gender balance in academic workforces, with some examples. We conclude by discussing collaboration with academic libraries and other stakeholders in developing dashboard tools and a community open knowledge coalition.

2. Rationale and Background

Leonelli et al. (2017, p. 194) discuss the notion of "*data shadows*" – data excluded fully or partially and the strategic or methodological reasons for absences. Our project critiques the selective use of data to stratify institutional performance by major world university rankings Academic World Ranking of Universities (AWRU), the Times Higher Education (THE) World University Ranking and Quacquarelli Symonds (QS). For example, they depend on publication and citation data extracted from commercial bibliographic databases Scopus (Elsevier) and Web of Science (Clarivate Analytics) whose coverage favours prestigious, English language, Global North publications with more emphasis on sciences, medicine and technology than the humanities and social sciences (Aksnes & Sivertsen, 2019; Tennant, 2020). Publication in these sources devalues diverse research from non-dominant and Indigenous populations, some countries and in disciplines such as the creative arts and political activist scholarship where journal publications are not the primary output format (Shahjahan & Wagner, 2019). This leads to a "geopolitics of knowledge" and a persistent coloniality (Stack & Mazawi, 2021, p. 226). These limited proxy measures with embedded epistemologies and questionable ontological assumptions distort the wider picture, and are open to institutional measures to influence outcomes and policy (McCormick, 2017).

The presence of rankings pressures university managements to develop strategies and practices in order for their institutions to compete on the ranking ladder and achieve status (Marginson, 2014). This occurs despite the limitations and exclusions of methodologies and the measurements rankings used as proxies for performance (Hazelkorn, 2018). For example, reputation and research performance, two key concepts of the AWRU, THE and QS (Selten et al., 2020) are built on citation analysis from the commercial databases discussed above. The use of "celebretization" measures such as Nobel Prize winners as a proxy of excellence and to rank reputation, perpetuates the dominance of white male demographics in universities (Stack, 2020, p. 5). These practices contradict the equity, diversity and inclusion (EDI) policies which many universities promote.

A key difference between our project and commercial datasets and analytical services is our approach to understanding scholarly communication from a critical research perspective. Our research emphasises the importance of asking how datasets might influence the questions asked about a university's performance, as well as the answers that are found. For example, a comparison of bibliographic data sources and their implications for university rankings finds significant differences in coverage of disciplines, publications and date ranges (Huang et al., 2020). Our project identifies and analyses public data sources to examine more deeply the outcomes of EDI policies and how this may affect research production, and institutional progress in achieving open access to knowledge. Such analysis can assist higher education institutions, authorities and researchers in understanding the extent and nature of their research performances, to become open to alternative ways of viewing research impact, to value openness in output and in research production. We aim to change the narrative about the role of research and knowledge production within institutions. The dataset takes a wide geographic scope and includes research publication data from countries in Asia, Africa and Latin America whose coverage by major commercial publishers and databases may be restricted, and across all disciplines. We engage with communities in Africa and Indigenous researchers in Australia to increase the visibility and accessibility of research within their communities, countries and the world.

2.1. Data Tools

The project develops interactive and visual dashboards in order to present multidimensional data narratives and analysis of research output

performance by institution, country and region. The interfaces include graphs, charts and numerical data as key communicative tools with which to present analysis (van Geenen & Wieringa, 2020). User-centred design represents data about research output in non-intimidating ways that enable researchers, administrators and executives to fill gaps in their knowledge of institutional research publishing practices and outcomes. Data analysis is presented through different visualisations representing a range of perspectives, revealing nuanced analysis of institutional and country research performance. In addition, we adapt, integrate and customise data analysis into reports, presentations, instructional materials, teaching and learning programmes. We build bespoke tools to assist institutions in understanding and assessing their actual and potential open research output, in negotiating publisher agreements, to inform and encourage staff and students to interrogate their research output data and performance metrics. Bringing together and providing access to the data used to assess, evaluate and characterise research output encourages critical inquiry of institutional and world university ranking processes, data and methods.

Donna Lanclos (2016) notes that knowing about and building on people's motivations and reasons to use tools and systems will encourage greater use. In the early stages of our project and to extend out of our data bubble, we sought the views of people from the communities we work with, who use and understand the data. We had shared the dashboard tool and data with senior executives and academic staff in our own institution, but wanted a wider institutional reach. To achieve this and move beyond anecdotal responses, we undertook research with library and research staff in Aotearoa New Zealand and Australian universities.

3. Methods and Data

In this section we discuss the project's dataset sources and analysis; the collection and analysis of diversity data; and methods used in the dashboard feedback survey. We use open data sources, including Microsoft Academic Graph, Unpaywall, Crossref metadata, Open Citations, ORCID and the Research Organization Registry (ROR) to analyse bibliographic, open access, affiliation, funder, publisher and citation data. The dataset includes more than 12 trillion items from over 140 countries and thousands of research institutions. Through our collection pipeline, we gather data about research publications

using unique Digital Object Identifiers (DOIs) from multiple sources, synthesise the datasets and create research performance calculations, including open access for each country and institution (see Data Availability statement).

To identify higher education institutional workforce demographic data relating to diversity we searched for publicly available regional and national web-based sources of statistics. Terminology varies and includes 'higher education', 'tertiary education', 'university/universities', 'ministry', 'department of education', 'statistics', 'gender', 'ethnicity', 'race' and 'disability' (Wilson et al., 2022a). Gender statistics are the most widely available, although primarily binary (men/women, male/female). We acknowledge limitations in how data are collected and reported, and identify the need for further exploration of the intersectionality of gender or sex and ethnicity, race or indigeneity, age and disability in understanding equity and diversity (Harris & Patton, 2019; Rice et al., 2019).

To gather feedback on the dashboard tool we undertook a survey in our region to obtain: (1) comments on the dashboard layout, navigation and the data incorporated therein; (2) insights into how institutions, library and research staff, researchers, executives and administrators might use the data, and associated behaviour around such use in the academic environment. In February 2020, we invited CAUL (Council of Australian University Librarians) and CONZUL (Council of New Zealand University Librarians) members via email to register to access the dashboard, review their institution's data, and to participate in an anonymous online Qualtrics survey, plus an optional follow-up telephone interview (Curtin University Human Research Ethics Committee approval HRE2020-0086). Twenty-seven universities out of 47 member institutions (39 CAUL and eight CONZUL) registered to access the dashboard. Five responses came from universities in Aotearoa New Zealand and 12 from Australia. Ten were fully completed, four partially and three minimally, with two follow up telephone interviews. Some more survey respondents indicated they would participate in a telephone interview, but as the responses were anonymous as specified in the ethics approval, we could not follow up.

The online survey included twelve questions:

- three related to institutional and work identity;
- nine related to the dashboard layout, navigation and documentation, the data analysis and its visual presentation.

Most questions were multiple choice and allowed for textual comments. One free text question asked respondents for additional comments and suggestions on the dashboard. The final survey question offered participants the option to expand on their responses in a follow up telephone interview with eight questions aimed to build qualitatively on the survey responses. They included two institutional questions, one related to the interviewees' work role, and five semi-structured questions focused on their responses to the usefulness and relevance of the data and dashboard (Wilson et al., 2022b).

In developing the dashboards, the project uses different visual software, including Google Data Studio and Kibana, an open source user interface for the ElasticSearch search analytics tool and web-based software.

4. Open Access Research Output

A key focus of the project is understanding the trajectory of open access institutional research output across the globe. The Budapest Open Access Initiative (BOAI, 2002) produced the first formal statement on open access (OA) free to read, download and share, with two options for opening research output: self-archiving into repositories and publication in OA journals. Self-archived (Green) OA is research free to read through an institutional or disciplinary repository or a preprint server. Published (Gold) OA is free to read either with an open publisher (no fee to publish), or via a paid publisher (with article processing charges or APCs), and distributed with an open licence. These options have expanded as commercial publishers entered the OA landscape. Hybrid OA is free to read from a subscription-based publisher (with APCs) and distributed with an open licence. Gold/Hybrid and Green can overlap when a self-archived copy exists as well as a published version.

Different self-archiving options include institutional and disciplinary repositories, and pre-print servers such as ArXiv for physics, mathematics and computer science, a pioneering model followed by other disciplines. Publishing models include digital only publishers who are exclusively OA, both non-profit (e.g. the Public Library of Science (PLOS)) and for profit (e.g. BioMed Central), as well as existing originally paper-based publishers who provide an OA option, usually involving APCs (Tennant et al., 2016). With rising costs to researchers of publishing OA commercially, research organisations and universities began to cancel institutional journal subscriptions. This has led to transformative or 'read and publish' OA agreements between publishers

and libraries or library consortia in many countries and regions, combining the costs of OA publishing for institutional authors, who generally retain copyright under a Creative Commons licence, with costs for reading publications (Hinchliffe, 2020).

Many OA publishing policies, statements and funder mandates are in place across the globe. Policies by country or region include Europe's Plan S policy requiring publicly and privately funded research from 2021 to be immediately available via OA journals, platforms or repositories (cOAlition S, n.d.). Institutional mandates or policies may specify OA through local repositories (Green) or by publisher (Gold); and funder mandates require research they fund to be available as open access. Publisher Gold OA fees can be prohibitive and out of reach for many researchers and regions, producing global inequities. The option of repository-based OA (Green) is cost free to authors, although it may be more complex to negotiate (Björk, 2017). In some parts of the world challenges with reliable Internet access, technology infrastructure, policies, expertise and funding impact the development and maintenance of institutional repositories (Wilson et al., 2020). However, our OA analysis highlights strong performance and initiatives at some regional and institutional levels.

Our OA analysis from 2010 to 2020 across six regions (Asia, Europe, North America, Latin America, Africa, Oceania) indicates the growth and shifts in Green (Repository) and Gold (Publisher) output over the ten-year period. These numbers reflect developing national policies and programs. In 2010, Latin American institutions (brown) appear with the highest Gold and Green output (Figure 1). Brazil and Chile developed the Scientific Electronic Library Online (SciELO) journal network in 1998 as an open platform for Latin American research output, to provide greater visibility of research publication than the major commercial databases such as Web of Science (Packer, 2020). An additional 13 countries, including Portugal, Spain and South Africa, have joined the network.

Figure 2 shows Latin American institutional research output moving further towards the Gold OA axis by 2020. The growth since 2010 of European institutions (green), clustering between 35% and 60% on the Green and Gold OA axes, reflects EU and UK policies that promote OA publishing, although these vary by country and institution. A small number of institutions in Asia (orange) reaches towards 80% Gold OA and up to 60% Green

Fig. 1: Publisher-mediated OA (Gold) vs repository-mediated OA (Green) research output by institution since 2010. Each point plotted is a university, with size indicating the number of outputs. Colour indicates the region. Data sources: Crossref, Unpaywall, Microsoft Academic. Analysis and image: Curtin Open Knowledge Initiative.

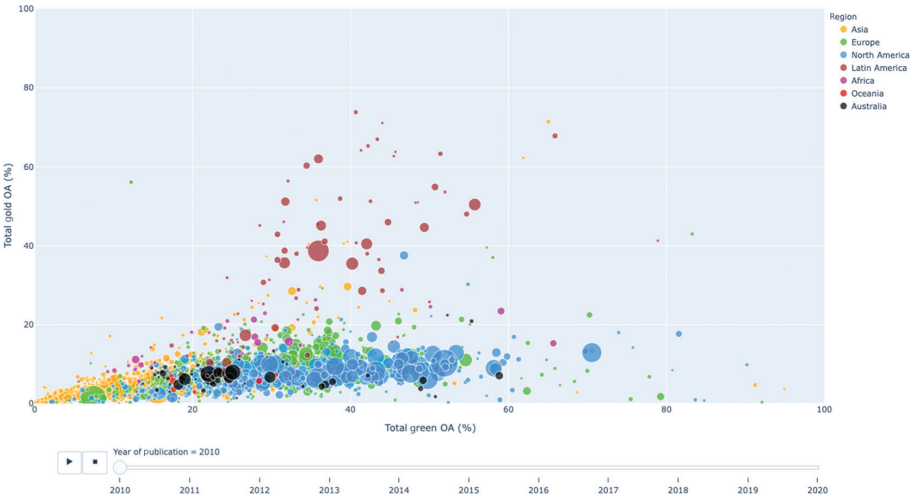


Fig. 2: Publisher-mediated OA (Gold) vs. repository-mediated OA (Green) by institution for 2020. Each point plotted is a university, with size indicating the number of outputs. Colour indicates the region. Data sources: Crossref, Unpaywall, Microsoft Academic. Analysis: Curtin Open Knowledge Initiative.

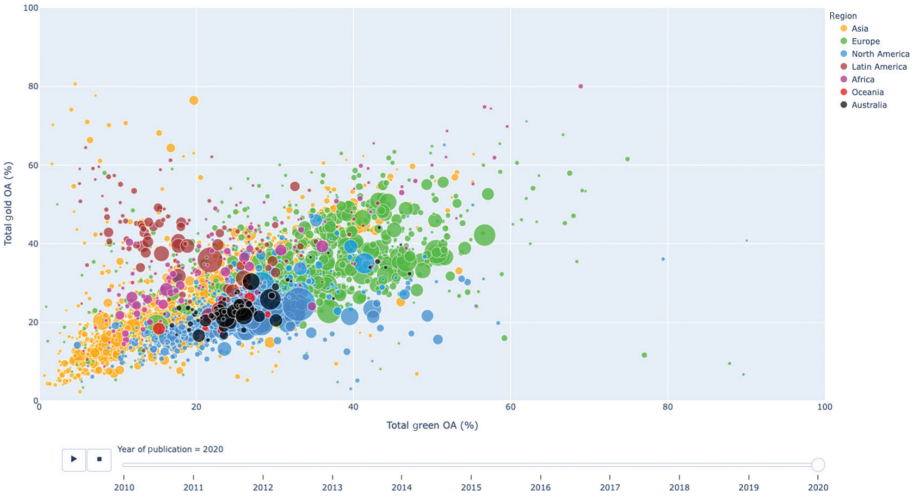
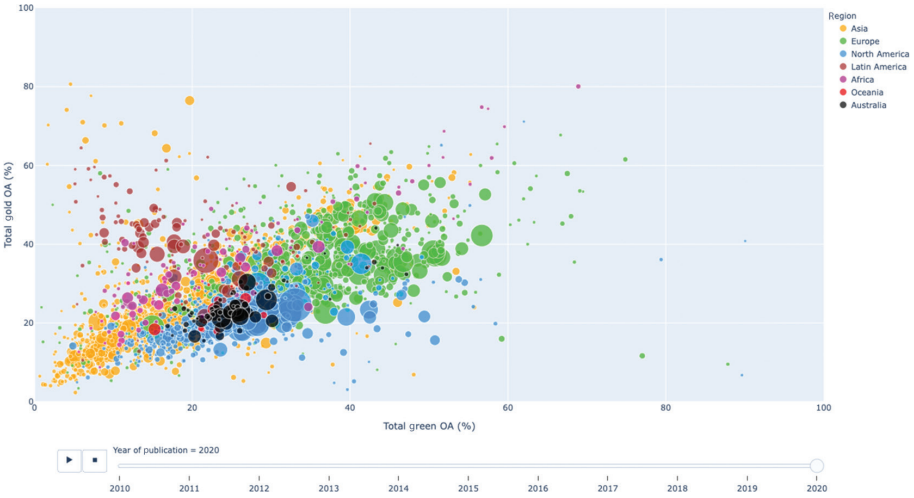


Fig. 3: The Animated version of OA Gold and Green trajectories for the years 2010–2020.

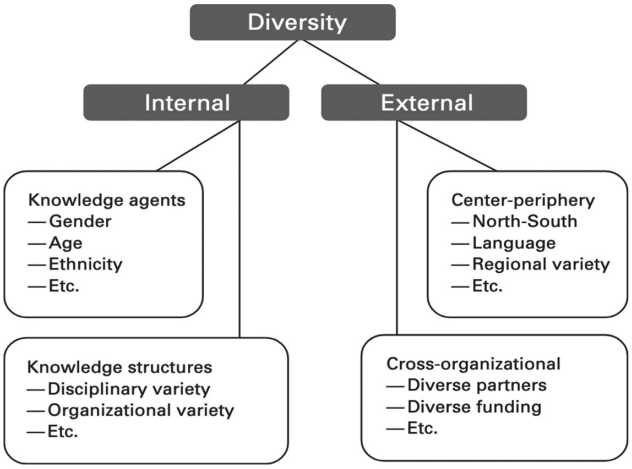


OA. Some institutions in Africa (pink) show high Green OA and Gold OA output. United States institutions (blue) lean towards Green OA with some reaching higher Gold OA levels, reflecting variations by institution and funder, and suggesting variant policies and practices. Australian institutions (black) hover between 20% and 40% Green OA and Gold OA, reflecting a lack of national policy although the two major public research funders have OA mandates. Figure 3 shows the animated version of OA Gold and Green trajectories for the years 2010–2020.

5. Diversity in Research Production

The Curtin Open Knowledge Initiative encourages universities to centre the principles of openness, to embrace digital open access, and to build diversity, equity and inclusion within institutions and in society (see Figure 4). Internal diversity includes demographic dimensions of researchers, teaching, management and administrative workforces, with diverse researchers as knowledge agents engaged across disciplines and communities. Structural diversity within institutions is achieved through equitable policies and practices and the sharing of data and knowledge. External diversity encompasses diverse research perspectives and knowledge sources, languages, geographical collaborations at regional, national and international levels, as well as diversity in funding sources (Montgomery et al., 2021).

Fig. 4: Dimensions of diversity in open knowledge institutions. Source: Montgomery et al. (2021).



Greater diversity in research workforces challenges structural and scholarly inequalities in the topics and perspectives investigated through institutional research, and awareness of the gaps in knowledge output and research funding is fundamental to addressing inequalities (Kozlowski et al., 2022). In relation to internal diversity, our project analyses public institutional demographic data in order to understand *who* is producing research and *how* knowledge is shared within and across disciplines, between universities and wider communities (Montgomery et al., 2021). We discuss gender, the public higher education data the most widely available.

5.1. Gender in Academic Workforces

Many higher education institutions have policies committed to hiring the best regardless of gender, but in reality, women make up only 20 to 30 percent of senior management in most universities around the world (Manfredi et al., 2019; Shepherd, 2017). A growing body of research documents persistent gender differences and gender asymmetry in research performance and productivity as measured by publication output, citations and funding grants in academic and research institutions (Holman et al., 2018; Huang et al., 2020). Analysis of recruitment practices in scientific disciplines in Italian universities

demonstrated no direct discrimination favouring one gender, but persistent "underrepresentation of women in the Italian academic systems" indicates gender inequality is present at a deeper level (Abramo et al., 2016, p. 137). Steinþórsdóttir et al. (2019, p. 3) analysed budgeting and management processes in a university in Iceland, revealing a "gendered academia" where an entrenched level of bias in assessment and bibliometric performance indicators favour the STEM disciplines and publications in elite journals. This dictates academics' salaries, funding promotion and status and perpetuates a lack of gender equality across the university. Viviane Albenga (2016) found low proportions of women in senior academic positions in one French university, highlighting differences between management perspectives on gender equality, such as parity on boards, and academic research knowledge. Research by Sá et al. (2020) among elite scientists reveals multiple factors that contribute to gender differences, including recognition, career length. Identifying lower numbers of citations for women as department chairs than men, they question the reliance on bibliometric indicators of research excellence such as citation analysis. Stack (2021) argues that the metrics and data used by university world rankings are incompatible with institutional inclusivity and equity particularly in the United Kingdom and North America where many universities continue to be dominated by policies and practices that favour white men, particularly in leadership positions. The dominance of male standards of excellence suggests institutional change involves structural and cultural repositioning.

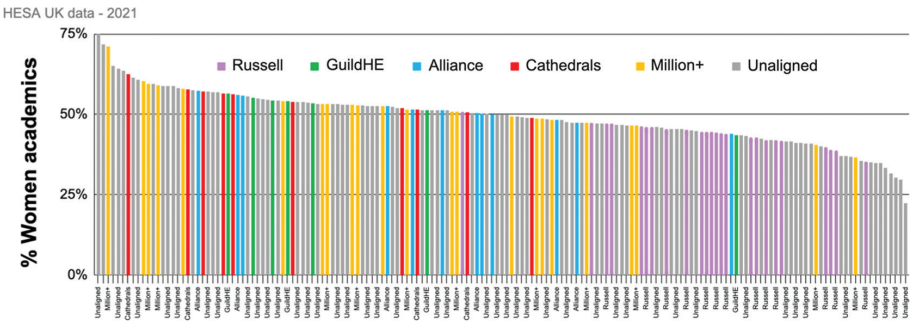
Gender equality data now feature in two world university rankings, although with some methodological limitations. The CWTS Leiden University Ranking of gender diversity for 1,225 research intensive universities provides proportions of authorship by men and women. Bibliometric analysis is based only on English language articles and reviews indexed in the Web of Science bibliographic database (Clarivate Analytics) from 2016 to 2019, and that appear in a core set of international journals CWTS identifies. To assign gender CWTS applies its algorithm for author name disambiguation based on author first names and author country, with approximately 70% success rate (Centre for Science and Technology, 2022).

The *Times Higher Education* impact Rankings 2022 assess the performance of 938 universities against the United Nations Sustainable Development Goal 5: Gender Equality. Twelve indicators include published gender equality research and papers in the top 10 percent of cited journals, the number of

publications, and percentages of research authored by women (based on Elsevier *Scopus* bibliographic data 2016–2020, with limitations in research coverage). The Elsevier methodology to determine author gender is not specified, and Bothwell et al. (2022) acknowledge the difficulties in assessing non-gendered names in some countries. Data and evidence submitted voluntarily by universities includes the proportion of senior women academics and policies and practices supporting the promotion and recruitment of women staff and students (Times Higher Education, 2022). Bothwell et al. (2022) note the limitations in obtaining such data from some institutions, and that the countries performing well are those with "abundant available documentation of their activities in the area of gender equality" (p. 7).

Llorens et al. (2021, p. 2067) note a lack of "systematic gathering and reporting" and sharing of gender data by institutions and recruiting committees. Analysis of gender workforce data is key to identifying and acting on gender disparities. Where data are available, we investigate institutional gender percentages in the context of research and reputational profiles of higher education workforces. As an example, in Figure 5 we present percentages of women academic non-atypical (on open-ended/permanent and fixed-term contracts) in United Kingdom higher education institutions in 2021 together with grouping data. The analysis uses sex data (the terminology adopted by HESA, the source agency) provided by institutions. Reporting academic staff data is required but since 2019/2020 not for non-academic staff except

Fig. 5: Percentages of women in the non-atypical academic workforce in a subset of higher education institutions in the United Kingdom by institutional groupings, 2021. Data source: Higher Education Statistics Agency (2022). Analysis and image: Curtin Open Knowledge Initiative. The full list of institutions and data is available at Zenodo https://doi.org/10.5281/zenodo.6629671.



"vice-chancellors/heads of institutions or governors which is mandatory" (Higher Education Statistics Agency, 2022). A subset of 2021 data for 163 institutions shows those with more than 50 percent women academics are mostly in the groupings Unaligned, University Alliance (professional and technical universities), Cathedrals (originally teacher training colleges), GuildHE (smaller, specialist, often creative arts institutions) and Million+(post-1992 universities). Only one university in the more prestigious Russell Group, with research intensive disciplines such as the sciences, technology, engineering, mathematics and medicine (STEMM) often dominated by men, employs more than 50 percent women academics.

The Athena SWAN Charter was introduced in the United Kingdom in 2005 to advance gender equality in academia, recognising inequalities in the STEMM disciplines and lower proportions of women in senior positions across universities (Advance HE, n.d). Despite improvements in the gender diversity of senior academics as a result of Athena SWAN (Xiao et al., 2020), our analysis in Figure 5 indicates gender disparities in academic staff in research intensive universities. Women's progression is often compounded by intersectional discrimination including race and ethnicity, sexual orientation and disability (Llorens et al., 2021).

Although many universities have diversity, equity and inclusion policies in place, when sharing analysis we have found limited consideration of the intersection and impacts of gender and other demographic diversity factors in research production. We investigate further correlations between gender and open access research output using indicators such as percentages of academic women, open access output and contextual factors of institutional reputation and location, and income (Huang et al., 2021; Wilson et al., 2022c). Sharing such analysis, we aim to open a conversation around alternative ways and indicators to view and understand connections between diversity and research.

6. Working with Academic Libraries

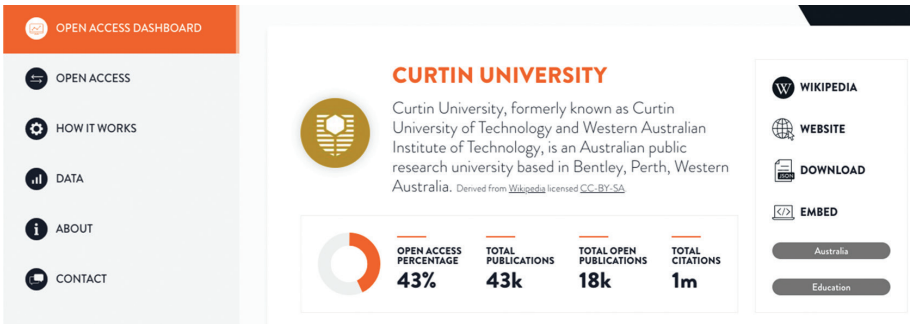
The Curtin Open Knowledge Initiative shares the dataset, analysis and visual dashboards with scholarly communication communities. We seek feedback and engagement from those who understand and use the data to make their own analyses such as academic libraries and research offices. They produce

and disseminate information about publishing and data sharing, promote open knowledge and skill development through the use of analytical tools, reporting and educational functions, advise and assist researchers and are well-placed to review data analysis and tools.

In 2020, we engaged Council of Australian University Librarians (CAUL) and Council of New Zealand University Librarians (CONZUL) members in a survey to review the dashboard interface and the relevance of the data analysis within their working environments. Respondents indicated the dashboard provided new perspectives on existing data, and access to new and previously unavailable data. In particular, they noted the value of institutional level OA publication analysis, compliance with research funder OA mandates and OA citation advantages which were difficult to obtain from other sources, including institutional repositories. Respondents indicated the analysis of OA citation figures would help researchers to appreciate the value of OA publishing and how it can benefit them, beyond simply following institutional policy imperatives to self-archive research. They expressed interest in access to more analysis, including underlying data and the OA performance of other institutions for comparative and benchmarking purposes. Overall, they indicated the data and its understanding was strengthened by being brought together in one location in a visual and interactive format (Wilson et al., 2022b).

This feedback provided valuable insights into the dashboard format and layout and researcher scholarly publishing practices and attitudes. First, the cognitive advantages of the visual dashboard interface indicate the strength of this method in communicating complex data and related narratives. Second, the participants' positive responses to the data content and dashboard, its accessibility and format encouraged us to continue building tools. Third, the interest expressed in comparative data contributed to our development of a public Open Access Dashboard (https://open.coki.ac). This dashboard provides publication, OA and citation data for 142 countries and 5,117 institutions with at least 1,000 research outputs. Users can search and display analysis by institution and country. Full details of data collection, the dataset code and data sources are available (Hosking et al., 2022). Figure 6 provides an example of an institutional dashboard entry with research performance analysis of total publications, the OA percentage and total citations.

Fig. 6: A summary of research performance for Curtin University, 2000–2021. Data sources: Crossref, Unpaywall, Microsoft Academic. Analysis and Image: Curtin Open Knowledge Initiative.



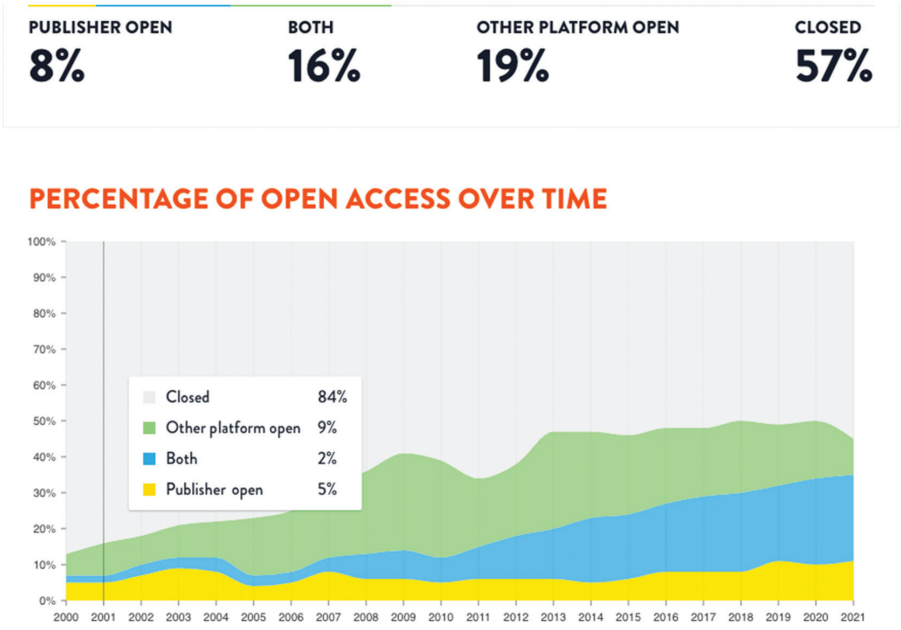
In Figure 7, the institutional OA percentage is broken down into 'Closed', 'Publisher Open' (in an OA Journal, sometimes referred to as Gold OA, Hybrid) and 'No Guarantees', (or Bronze); 'Other Platform Open' (deposited in an institutional repository, a disciplinary repository, a preprint repository or a general repository, sometimes referred to as Green OA), and 'Both' ('Publisher Open' and 'Other Platform Open'). 'Both' indicates an item may be published open simultaneously with a copy in a repository, representing two pathways for creating open and accessible research. We introduce new terminology to clarify the sometimes-confusing colour labelling for OA types (such as Green, Gold, Hybrid, Bronze).

The insert box data in Figure 7 provides details of research output in the year 2001. The longitudinal data indicate changes over the two decades. These data and images are available for all institutions and countries included in the dataset at present are those with over 1,000 research outputs. Development of the dashboard continues, adding more search features.

7. Conclusion: Sharing and Moving Forward

Working from a critical research perspective the Curtin Open Knowledge Initiative project has built a dataset and created dashboard analytical tools that provide new insights into the research performance of higher education institutions. Using our data and analysis, institutions can explore how open

Fig. 7: Types of open access research output for Curtin University, 2000–2021. Data sources: Crossref, Unpaywall, Microsoft Academic. Analysis and Image: Curtin Open Knowledge Initiative.



are the knowledges they produce and how their workforces reflect the diversity of student bodies, communities and wider populations. The analysis can provide indications of diverse research perspectives that may be lacking as a result of employment practices and preferences. The data analysis reveals complex national and regional patterns of research performance across the world, in particular OA output.

The project's goals are to deliver accessible, shareable data, resources, analysis and to open conversations about ways to open knowledge. Contributions and feedback received from academic libraries, researchers, funders and publishers in the scholarly communication community validate and confirm the need for accessible and neutral data that form the basis and origins of this project. We regard information as a community asset, and highlight an opportunity to move beyond commercial data towards the construction of resources that are governed by the higher education and research community. Collaboratively we have built a community coalition to meet a vital

need for maintaining access to complex research performance analysis and extending open knowledge production. The coalition includes stakeholders from higher education consortia, funders, research managers, libraries and publishers. Central components of the Open Knowledge community coalition include building and maintaining the data asset we have developed, a community of practice around making change, sharing expertise, software and data on evaluation and openness. Through sharing data and dashboard code with coalition members the Open Knowledge project expands its scope and analysis of institutional openness and knowledge making, diversity and inclusion in knowledge production, and evaluation of open knowledge performance.

Data availability

The COKI project dataset is available at Zenodo: https://doi.org/10.5281/zenodo.6399463.

References

Abramo, G., D'Angelo, C. A., & Rosati, F. (2016). Gender bias in academic recruitment. *Scientometrics*, *106*(1), 119–141. https://doi.org/10.1007/s11192-015-1783-3

Advance HE. (n.d). *Athena SWAN Charter*. Retrieved August 17, 2022, from https://www.ecu.ac.uk/equality-charters/athena-swan/about-athena-swan/

Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of scopus and web of science. *Journal of Data and Information Science*, *4*(1), 1–21. https://doi.org/10.2478/jdis-2019-0001

Albenga, V. (2016). Between knowledge and power: Triggering structural change for gender equality from inside in higher education institutions. In M. Bustelo, L. Ferguson, & M. Forest (Eds.), *The politics of feminist knowledge transfer: gender training and gender expertise* (pp. 139–156). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-48685-1_8

Björk, B.-C. (2017). Gold, green, and black open access. *Learned Publishing*, *30*(2), 173–175. https://doi.org/10.1002/leap.1096

BOAI. (2002, February 14). *Budapest Open Access Initiative*. https://www.budapestopenaccessinitiative.org/read

Bothwell, E., Roser Chinchilla, J. F., Deraze, E., Ellis, R., Galán-Muros, V., Gallegos, G., & Mutize, T. (2022). *Gender equality: How global universities are performing Part 1.* Times Higher Education, UNESCO International Institute of Higher Education in Latin America and the Caribbean (IESALC). https://www.timeshighereducation.com/sites/default/files/the_gender_equality_report_part_1.pdf

Centre for Science and Technology. (2022, June 22). *CWTS Leiden Ranking 2022*. http://www.leidenranking.com

cOAlition S. (n.d.). *Plan S Principles and Implementation*. Plan S. Retrieved 13 May 2022, from https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/

Gadd, E. (2021). Mis-measuring our universities: Why global university rankings don't add up. *Frontiers in Research Metrics and Analytics, 6*, Article 680023. https://doi.org/10.3389/frma.2021.680023

Harris, J. C., & Patton, L. D. (2019). Un/Doing intersectionality through higher education research. *The Journal of Higher Education*, *90*(3), 347–372. https://doi.org/10.1080/00221546.2018.1536936

Hazelkorn, E. (2018). Reshaping the world order of higher education: The role and impact of rankings on national and global systems. *Policy Reviews in Higher Education*, *2*(1), 4–31. https://doi.org/10.1080/23322969.2018.1424562

Higher Education Statistics Agency. (2022, February 17). *Who's working in HE?: Personal characteristics*. HESA. https://www.hesa.ac.uk/data-and-analysis/staff/working-in-he/characteristics

Hinchliffe, L. J. (2020, February 6). *Revisiting — Transformative Agreements: A Primer*. The scholarly kitchen. https://scholarlykitchen.sspnet.org/2020/02/06/revisiting-transformative-agreements-a-primer/

Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology*, *16*(4), Article e2004956. https://doi.org/10.1371/journal.pbio.2004956

Hosking, R., Diprose, J. P., Roelofs, A., Chien, T.-Y., Montgomery, L., & Neylon, C. (2022). *COKI Open Access Dataset* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6815932

Huang, C.-K. (Karl), Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020a). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, *1*(2), 445–478. https://doi.org/10.1162/qss_a_00031

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020b). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, *117*(9), 4609–4616. https://doi.org/10.1073/pnas.1914221117

Huang, C.-K. (Karl), Wilson, K., Neylon, C., Ozaygen, A., Montgomery, L., & Hosking, R. (2021). Mapping open knowledge institutions: An exploratory analysis of Australian universities. *PeerJ*, *9*, e11391. https://doi.org/10.7717/peerj.11391

Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences 119*(2), Article e2113067119. https://doi.org/10.1073/pnas.2113067119

Lanclos, D. (2016). Ethnographic approaches to the practices of scholarly communication: Tackling the mess of academia. *Insights*, *29*(3), 239–248. https://doi.org/10.1629/uksg.316

Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. *Science, Technology, & Human Values, 42*(2), 191–202. https://doi.org/10.1177/0162243916687039

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W. Y., Kidd, C., Marcelle, E., … Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron*, *109(*13), 2047–2074. https://doi.org/10.1016/j.neuron.2021.06.002

Manfredi, S., Clayton-Hathway, K., & Cousens, E. (2019). Increasing gender diversity in higher education leadership: The role of executive search firms. *Social Sciences, 8*(6), Article 168. https://doi.org/10.3390/socsci8060168

Marginson, S. (2014). University rankings and social science. *European Journal of Education*, *49*(1), 45–59. https://doi.org/10.1111/ejed.12061

McCormick, A. C. (2017). The intersection of rankings with university quality, public accountability and institutional improvement. In E. Hazelkorn (Ed.), *Global rankings and the geopolitics of higher education* (pp. 205–215). Routledge.

Montgomery, L., Hartley, J., Neylon, C., Gillies, M., Gray, E., Herrmann-Pillath, C., Huang, C.-K., Leach, J., Potts, J., Ren, X., Skinner, K., Sugimoto, C. R., & Wilson, K. (2021). *Open Knowledge Institutions: Reinventing Universities*. MIT Press. https://doi.org/10.7551/mitpress/13614.001.0001

Packer, A. L. (2020). The pasts, presents, and futures of SciELO. In M.-P. Eve & J. Gray (Eds.), *Reassembling scholarly communications histories, infrastructures, and global politics of open access* (pp. 297–313). MIT Press Direct. https://doi.org/10.7551/mitpress/11885.003.0030

Rice, C., Harrison, E., & Friedman, M. (2019). Doing justice to intersectionality in research. *Cultural Studies ↔ Critical Methodologies*, *19*(6), 409–420. https://doi.org/10.1177/1532708619829779

Sá, C., Cowley, S., Martinez, M., Kachynska, N., & Sabzalieva, E. (2020). Gender gaps in research productivity and recognition among elite scientists in the U.S., Canada, and South Africa. *PLoS One*, *15*(10), Article e0240903. https://doi.org/10.1371/journal.pone.0240903

Selten, F., Neylon, C., Huang, C.-K., & Groth, P. (2020). A longitudinal analysis of university rankings. *Quantitative Science Studies*, *1*(3), 1109–1135. https://doi.org/10.1162/qss_a_00052

Shahjahan, R. A., & Wagner, A. E. (2019). Unpacking ontological security: A decolonial reading of scholarly impact. *Educational Philosophy and Theory*, *51*(8), 779–791. https://doi.org/10.1080/00131857.2018.1454308

Shepherd, S. (2017). Why are there so few female leaders in higher education: A case of structure or agency? *Management in Education*, *31*(2), 82–87. https://doi.org/10.1177/0892020617696631

Stack, M. (2020). Academic stars and university rankings in higher education: Impacts on policy and practice. *Policy Reviews in Higher Education*, *4*(1), 4–24. https://doi.org/10.1080/23322969.2019.1667859

Stack, M., & Mazawi, A. E. (2021). Beyond rankings and impact factors. In M. Stack (Ed.), *Global University Rankings and the politics of knowledge* (pp. 225–242). University of Toronto Press.

Steinþórsdóttir, F. S., Pétursdóttir, G. M., Einarsdóttir, Þ., & Feuvre, N. L. (2019). New managerialism in the academy: Gender bias and precarity. *Gender, Work and Organization*, *26*(2), 124–139. http://dx.doi.org/10.1111/gwao.12286

Tennant, J. P. (2020). Web of Science and Scopus are not global databases of knowledge. *European Science Editing*, *46*, Article e51987. https://doi.org/10.3897/ese.2020.e51987

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, Chris. H. J. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review. [version 3; peer review: 4 approved, 1 approved with reservations]. *F1000Research*, *5*(632), 1–3. https://doi.org/10.12688/f1000research.8460.3

Times Higher Education. (2022, April 18). *Impact Rankings 2022: Gender equality (SDG 5) methodology*. https://www.timeshighereducation.com/impact-rankings-2022-gender-equality-sdg-5-methodology

van Geenen, D., & Wieringa, M. (2020). Approaching data visualizations as interfaces: An empirical demonstration of how data are imag(in)ed. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 141–156). Amsterdam University Press. https://doi.org/10.2307/j.ctvzgb8c7.15

Wilson, K., Kiuna, A., Lamptey, R., Veldsman, S., Montgomery, L., Neylon, C., Hosking, R., Huang, K., & Ozaygen, A., (2020). Open access and research dissemination in Africa. [Conference communication]. HAL sciences ouverte. http://doi.org/10.4000/proceedings.elpub.2020.20

Wilson, K., Neylon, C., Montgomery, L., Huang, C.-K. (Karl), Handcock, R. N., Roelofs, A., Hosking, R., & Ozaygen, A. (2022a). Global diversity in higher education workforces: Towards Openness. *Open Library of Humanities*, *8*(1), Article 1. https://doi.org/10.16995/olh.4809

Wilson, K., Montgomery, L., Neylon, C., Huang, C.-K. (Karl), Hosking, R., Handcock, R. N., & Ozaygen, A. (2022b). *Curtin Open Knowledge Initiative CAUL/CONZUL Institutional Dashboard Feedback Report, May 2020*. [Report]. Zenodo. https://doi.org/10.5281/zenodo.6537591

Wilson, K., Huang, C.-K. (Karl), Montgomery, L., Neylon, C., Handcock, R. N., Ozaygen, A., & Roelofs, A. (2022c). Changing the academic gender narrative through open access. *Publications*, *10*(3), 22. https://doi.org/10.3390/publications10030022

Xiao, Y., Pinkney, E., Au, T. K. F., & Yip, P. S. F. (2020). Athena SWAN and gender diversity: A UK-based retrospective cohort study. *BMJ Open*, *10*, Article e032915. https://doi.org/10.1136/bmjopen-2019-032915