Universities as Open Knowledge Institutions: Sharing vital research

Katie Wilson
Curtin University, katie.wilson@curtin.edu.au

Lucy Montgomery

Cameron Neylon

Richard Hosking

Chun-Kai (Karl) Huang

See next page for additional authors

https://docs.lib.purdue.edu/iatul/2021/wednesday/2

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Presenter Information
Katie Wilson, Lucy Montgomery, Cameron Neylon, Richard Hosking, Chun-Kai (Karl) Huang, Rebecca N. Handcock, Alkim Ozaygen, and Aniek Roelofs

This event is available at Purdue e-Pubs: https://docs.lib.purdue.edu/iatul/2021/wednesday/2
Universities as Open Knowledge Institutions: Sharing vital research

Katie Wilson, Lucy Montgomery, Cameron Neylon, Richard Hosking, Chun-Kai (Karl) Huang, Rebecca N. Handcock, Alkim Ozaygen, Aniek Roelofs

Curtin University, Australia
katie.wilson@curtin.edu.au

Abstract

Universities are key creators of knowledge. Ensuring that research outputs are not inaccessible behind paywalls, and that research data can be interrogated and built upon is central to efforts to improve the effectiveness of global research landscapes. Mandating and promoting open science and open access (OA) for published research outputs and sharing research data are important elements of building a open knowledge system, but there are additional benefits. Supporting diversity within knowledge-making institutions; enabling collaboration between universities and communities; addressing inequalities in access to knowledge resources and opportunities for contributing to knowledge making are also important. New tools are needed to help universities, funders and communities understand the extent to which a university is operating as an effective open knowledge institution; as well as the steps that might be taken to improve open knowledge performance. The spread of the COVID-19 pandemic has demonstrated the vital need for open research and knowledge to help find a global solution.

The Curtin Open Knowledge Initiative (COKI) is a Curtin University funded strategic research project exploring ways of understanding and assessing institutional progress towards knowledge openness through analysis of research output, diversity data, policies and outcomes, and access to knowledge resources including libraries. This paper discusses the COKI team’s multidimensional approach to assessing institutional knowledge openness beyond, but including, measures of OA, open science and open data. The COKI dataset of more than 12 trillion items enables exploration and analysis of many questions around publication, impact, research performance, university engagement, diversity and access to knowledge. This information can assist universities, libraries, funders and communities to understand and enhance institutional open knowledge performance and contributions.

Keywords: Open knowledge; open science; open access; diversity; COVID-19

Introduction

In 2021 the world is in an emergency health crisis requiring immediate, ongoing open access to research knowledge and data relating to coronaviruses, the COroona VIrus Disease 2019 or COVID-19. In order to effectively respond to the COVID-19 global health emergency research into the virus’s epidemiology and public health implications is needed. This includes the investigation of patterns of transmission, socioeconomic impacts, quarantine guidelines and strategies for prevention. So too is genome sequencing, the development of new vaccines and drug therapies, knowledge of short and long term complications, nursing approaches, and much more. The research response demanded by COVID-19 is, by its nature, multidisciplinary. As such, helping a global community of researchers to respond quickly and effectively to COVID-19 is best achieved by opening all research to provide free access to output in all disciplines. Ensuring that research data, analysis and findings are openly available enables international collaboration and sharing of essential research among nations with differing levels of expertise, budgets and COVID-19
infections (Lee and Haupt, 2021). Previous public health emergencies such as those involving SARS and MERS human coronaviruses and the Zika virus identified the need for open research, data and publications. During the COVID-19 pandemic preprints are providing instant access to pre-peer-reviewed research. New protocols for sharing research data have been developed, promoted and mobilised. Commercially published academic research became open as major publishers, societies and funders agreed to share the data and results of academic endeavours (Wellcome Trust, 2020, January 31). While magnanimous in a time of crisis, this opening of research also demonstrates the unreasonableness of price tags on valuable knowledge within a scholarly publishing marketplace dominated by commercial players and reinforced by university world rankings and their influence on universities.

The Curtin Open Knowledge Initiative (COKI) promotes openness in research and produces data analysis in order to understand the progress of universities and research institutions towards open knowledge institutions. This includes analysis of research output by extent and type of open access, national and international collaboration, publisher and funder performance, disciplinary output, citation advantage and diversity in research and institutional workforces. In this paper we discuss the COKI project and our analysis of research output data in order to understand multiple dimensions of global research and institutional knowledge openness. We review the impact of the COVID-19 pandemic within the environment of scholarly communication and its relationship to global survival with an overview of the COVID-19 open research landscape.

The open knowledge initiative

The COKI research project grew from a critique of world university rankings and related publishing and assessment practices within institutions. Its focus is on understanding and promoting openness in research and knowledge production through analysis of global institutional research output. To build understanding of the extent and nature of scholarly research and openness, COKI gathers data about institutional research output in over 190 countries. Using multiple sources, the project captures data at large scale (via a 12 trillion plus dataset) producing analysis from different organisational and geographic perspectives. We aggregate and analyse bibliographic data, such as researcher affiliation, publication open access status, funder data, and citations from sources including Microsoft Academic, ORCID, CrossRef, Unpaywall, GRID, ROR, and OpenCitations. The project presents analysis by institution, country, consortia, funder, and publisher to deliver a unique data view for these different communities.

The project aims to encourage dialogue and facilitate understanding of institutional, consortial and national openness in research performance, open access, collaboration, and levels of open research by funders and publishers. We explore the progress of institutional open access policies on research outcomes (Huang et al., 2020). With this understanding of research performance and outcomes institutions can review assessment, evaluation practices, and attitudes that are driven by compliance with rankings, and focus on openness as an alternative means of assessing research impact. Cultural change at national and institutional systemic levels is central to achieving openness through communication and coordination of policies, actions and outcomes and diversity in workforces and research production (Montgomery et al., 2021).

Visualisation of research output analysis

COKI develops interactive visual dashboards built on the dataset as tools for understanding the details of scholarly communication institutions create through research and to encourage dialogue within and among institutions. Data visualisation through these dashboards enables stakeholders to understand and assess their performance and positioning within the scholarly communication environment, and to see opportunities for change.
Figure 1 below shows the front/landing page of COKI’s Curtin University Research Dashboard. This dashboard summarises Curtin University based publications, citations, unique funders and publishers across the years 2000 to 2020, and breaks down this data by the amount and type of open access (i.e., by publisher (gold), repository (green), or via a hybrid subscription journal with both OA and non-open publications).

Figure 1: Curtin University Research dashboard front/landing page (Analysis and images: COKI)

Web links from the front/landing page to further dashboard pages include analysis of national and international collaboration by Curtin academics (Figure 2), research output and open access analysis by discipline, publisher, funder (acknowledged in publications), and citation advantages provided by open access.
Figure 2: Curtin University collaborations with the United States of America, 2000-2020, mapping and top eight institutions (Analysis and images: COKI)

Institutional data analysis such as this makes possible conversations and correlations around scholarly output dimensions, for example, to assess which factors contribute to open access output and international collaboration.

COKI provides two public dashboards through its website (http://openknowledge.community/insights/). The Country Open Access dashboard displays for over 190 individual countries their publication research output data, broken down by non-open access status, open access type, and comparison of open and non-open article citation rates. The Research Funding dashboard shares analyses per country of funders acknowledged in publications, as well as analyses of the ratio of domestic funding to international funding per country. COKI creates research dashboards for groups of institutions, and dashboards focused on specific subject orientations, for example, Climate Change literature. Automated country reports and visual animations showing longitudinal analysis such as the evolution of publisher (gold) and repository (green) open access output (https://storage.googleapis.com/oaspa_talk_files/institution_scatter.html) enable the creation of data narratives to meet the needs of different communities and perspectives. COKI’s raw data, software and code are open source and available for others to use.

The visualisation and interactivity of the dashboards and animations facilitate understanding of scientometric and bibliometric data, significantly reducing the “cognitive load” in decision making (Chen, 2020, p. 12). We find this format enables the transmission and communication of data analysis to a range of institutional members including senior executives and researchers.

Diversity

Universities are key creators of research, but in order to support open knowledge production and mobilisation they must embrace opportunities to engage diverse researchers and communities in knowledge making and knowledge sharing. To understand and explore correlations between researchers and knowledge production COKI analyses public institutional workforce diversity dimensions of gender, origin and disability. We gather national statistics for individual universities and research organisations, where available. This enables understanding of who is involved in the creation of knowledge, how knowledge is shared across disciplines and to what extent university workforces reflect the diversity in their student bodies, local communities and population groups.
Data about gender/sex (women, men) are shared by most countries. Some collect a third gender category (unknown or unspecified) but for privacy reasons because of small numbers these data are often not published. Statistics relating to origin in workforces vary considerably and include ethnicity, race, nationality, citizenship, indigeneity, minorities. Collecting origin data does not occur in some countries, and past and present racial biases and prejudices create a reluctance for individuals to indicate status of origin. Similarly, staff members may not divulge disabilities because of concerns about impacts on careers and job security. Although many countries have enacted legislation to address workforce disability discrimination, only a few countries and institutions collect or make statistics public. The small numbers of people with acknowledged disabilities limits the inclusion of their perspectives in research knowledge (Wilson et al., 2020).

Where possible we incorporate diversity analysis into the visual dashboards alongside research performance data. Although these data are public, institutional members are not widely familiar with their own workforce diversity or consider it in relation to research production.

COVID-19 and open knowledge

The COVID-19 pandemic has brought into sharp focus the significance and vital importance of open access to science research, knowledge, expertise and data. This means immediate and real-time sharing of research results such as epidemiological studies, genomic data analysis, modelling and managing hospital occupancy, vaccine development and more, to enable and facilitate collaboration, medical research, and public health decision-making and policy development. These sharing actions define the logic and sense of such practices, with interdisciplinary calls for openness in research to continue beyond the current pandemic (Jamali, Barkemeyer, Leigh, and Samara, 2020). The Wellcome Trust notes the “shortcomings of the traditional scholarly publishing system, which is not fit for purpose in the 21st century” (Kiley, May 21, 2020). COVID-19 research undertaken at universities and research institutes is often funded publicly and used in the commercial development of drugs and vaccines. COVID Moonshot, open drug discovery research to build a vaccine pill, is an example of global collaborative research across disciplines, public and private organisations (Delft et al., 2021).

Preprint services BioRxiv and medRxiv began early sharing of COVID-19 research output, with the first BioRxiv preprint posted on 19 January 2020 (Chen et al., 2020a) published openly in Infectious Diseases of Poverty on 28 February 2020 (Chen et al., 2020b). At the time of writing, preprint services provide open access to 16,623 COVID-19 related preprints (12,804 medRxiv, 3,819 bioRxiv). In January, 2020, the Wellcome Trust requested researchers, journals, publishers and funders to make open their COVID-19 related research, invoking the 2016 Statement on data sharing in public health emergencies developed in response to the Ebola and Zika outbreaks (Wellcome Trust, January 31, 2020). Research bodies and more than 50 publishers opened their previously tolled doors to make published research output freely accessible with a Creative Commons CC-BY licence in special collections. PubMed Central (PMC) and other public repositories made research available at the time of publication. However, analysis in April 2020 by Arrizabalaga et al. (2020) of 5,611 COVID-19-related articles from PubMed found that although publications were open, 68.3% held the undefined Bronze OA licence and 72.1% had no specific licence regarding reuse. Publishers Springer, Elsevier and Wiley provided statements allowing temporary use for the duration of the pandemic (p. 4-5). Opening access to publications also highlights closed access to earlier research on coronavirus ranging back to 1988 that may be cited in these now open publications. In other COVID-19 global publishing activities, the Chinese government directed locally funded research to be published in Chinese, not international journals, limiting access to publications (Larivière, Shu, and Sugimoto, 2020, March 5).
The rapid growth in COVID-19 preprints and publications led to a “discoverability crisis” (Kraker, Schramm, and Kittel, 2021, p. 4), creating a chaotic “knowledge ecosystem” (Zhang et al., 2021, p. 4235). Expedient means to enhance discovery of relevant and recent material soon emerged. For example, LitCOVID (National Center for Biotechnology Information, 2021) includes PubMed articles with data downloadable in two formats, for text mining with annotations and as bibliographic citations. The World Health Organisation (WHO) (2021) database gathers international multilingual published scientific findings and knowledge on COVID-19 from multiple databases. The Lancet COVID-19 Resource Centre (https://www.thelancet.com/coronavirus) provides open access to all research, commentary, and analysis in the journal collection. Cell Press Coronavirus Resource Hub offers free access to content in Cell Press journals. The COVID-19 Open Research Dataset (CORD-19) (Semantic Scholar, n.d.) from the Allen Institute for AI offers tools to locate and visualise relevant data and research with networks of linkages between authors, institutions and topics. The COVID-19 Knowledge Graph, a non-profit collaborative project from academia and industry, integrates several COVID-19 datasets to bring together and visualise related COVID-19 publications, statistics, genetic and molecular data (CovidGraph Website, 2020). The COVID-19/SARS-CoV-2 papers from medRxiv, bioRxiv, ArXiv, Elsevier and PubMed using natural language processing to filter relevant material (Collabovid. n.d.). Kraker et al. (2021) identify infrastructure problems, characterizing a lack of innovation in commercial database and academic search services, and outline the promise of community based open discovery infrastructure. Their own collaborative project CoVis is a curated knowledge map of seminal works in eight biomedical areas, providing rapid access to relevant literature (CoVis, 2020).

Many questions related to openness in publishing remain in relation to COVID-19 research. For how long will COVID-19 publisher content remain open and to what extent? Some publishers and societies have committed to maintaining open access (e.g., The Royal Society, 2021). Others set end dates on their openness. For example, the Elsevier Coronavirus Research Hub licensed output exclusively for Coronavirus research ceased on 30 April 2021, although guidelines for clinicians and patients and selected research publications remain open (Elsevier, April 30, 2021). A search of the Dimensions database using the pre-set search string ["2019-nCoV" OR "COVID-19" OR “SARS-CoV-2” OR "HCov-2019" OR "hcov" OR "NC0V19" OR "severe acute respiratory syndrome coronavirus 2" OR "severe acute respiratory syndrome corona virus 2" OR “coronavirus disease 2019” OR (("coronavirus" OR "corona virus") AND (Wuhan OR China OR novel)))] found 492,544 publications dated 2020-2021. Of these, 72.73 % (358,257) are open via gold, bronze (no reuse licence, approximately 20%), green and hybrid routes, calculated using Unpaywall data (Digital Science & Research Solutions, 2021, June 17). By comparison, only 38.2 per cent (784,371) of 2,052,797 publications in Dimensions with terms HIV or HIV/AIDS or HIV-1 are identified as open access. WHO classifies HIV/AIDS as an epidemic while some researchers characterise it as a pandemic in countries such as South Africa and localised in countries such as the United States (Eisinger and Fauci, 2018). Will COVID-19 research retract to a similarly low level of openness? As new COVID-19 strains develop the need for research continues and sharing knowledge outcomes of vaccines globally continues to be vital.

What are the outcomes of the global opening of COVID-19 research and data on research practice and the science that arises from it? The Wellcome Trust, UKRI and the Bill and Melinda Gates Foundation (2021, May 24) commissioned research to understand the extent to which commitments made by more than 160 organisations who signed the Wellcome Trust January 2020 statement were put into practice, to identify and assess the impacts on policies and practices, researcher behaviour, including responses from WHO and other public health agencies.

How does opening up the COVID-19 literature influence thinking and practices around open knowledge? Hayashi (2021) discusses how COVID-19 has changed the scholarly publishing environment by extending and re framing existing open science/open access practices such as preprints, open peer review, and identifies the need for rapid, expedited preprint review processes. The systems and technologies are already in place for such change. Moving from “publish or perish” to “share or perish” (Hayashi, 2021, p. 2) succinctly encapsulates the change needed: the
emergency of COVID-19 has shown why open access is essential. Commercial publishers have opened doors, but for how long? Kraker et al. (2021) suggest what is needed for open infrastructure to survive: align tools with use cases; extend to countries, languages and disciplines beyond the dominant western discovery frameworks; combine AI and “human intervention”; and adopt alternative funding models to sustain open discovery (pp. 10-12).

Geopolitical differences and tensions

COVID-19 reveals underlying geopolitical research tensions. Lee and Haupt (2021) compare pre-COVID-19 output (2015 to 2019) with a set of 3401 COVID-19 related publications from 111 countries dated 12 January to 9 May 2020 from the Scopus citation indexing database. They find a positive correlation between the impact of COVID-19 and a country’s participation in international collaboration and publishing open access. However, national Gross Domestic Product (GDP) affects these publishing patterns: wealthier countries with high GDP (i.e., the United States and China in this analysis) tend to collaborate less and publish less open access. COKI’s public dashboards show the highest ratios of domestic to international funders in China (624.39%) and the USA (188.04%), and lower open access output in 2020: China (34.02%), the USA (43.76%). Better resourced and with high COVID-19 case numbers, Lee and Haupt (2021) argue, researchers in these countries are well-equipped to conduct their research domestically, practising “scientific nationalism” as opposed to “scientific globalism” (p. 953). The small dataset is a limitation to this analysis, and the use of the Scopus dataset excludes many publication sources and languages beyond Europe and North America, from Africa, South America and some of Asia (Tennant, 2020). However, the study provides insights into national and international imperatives to sharing and globalising research. Others attribute a slowdown in collaborations between China and the USA to Chinese government restrictions on sharing COVID-19 information and scrutiny of such collaborative research in the USA (Cai, Fry and Wagner, 2021; Maher and Noorden, 2021).

Mencía-Ripley, Paulino-Ramírez, Jiménez and Camilo (2021) draw attention to ongoing practices of colonisation within scientific research structures. “Science diplomacy” where academic researchers participate in national programs and policies that have an international reach (p. 1) is dominated by developed nations, continuing colonial domination. COVID-19 affects more nations, populations and communities in the Global South, but their needs and voices are heard less in research and development. Mencía-Ripley et al. (2021) encourage greater South-South collaboration and less reliance on North-South relationships and funding. The Tropical Medicine and Global Health Institute at the Universidad Iberoamericana in the Dominican Republic, a low-middle income country, contributes to national diagnosis of SARS-CoV-2 infections and local genome sequencing. Universities have a role in increasing South-South collaborative research, contributing to “technology sovereignty” and policy change (p. 3).

COVID-19 research and diversity

Diversity disparities in academic workforces vary globally and affect the production of research and knowledge through uneven representation. Research into the COVID-19 pandemic crisis lockdowns and academic work practices shows impacts on research productivity for women researchers. Lockdowns and remote working produced higher levels of home and child care duties and financial insecurities for women scientists and researchers. Minority populations are vulnerable to COVID-19 in some countries (Chakraborty, 2020). Research with, data, and funding for non-white populations and researchers are less plentiful (Crooks, Donenberg, and Matthews, 2021). Analysis highlights disparities and reinforces the need for awareness and addressing higher education workforce diversity and equity to benefit diverse populations affected by the pandemic (Australian Academy of Science, 2020; Esser et al., 2020; King and Frederickson, 2020; Woolston, 2020).
COKI and COVID-19

The COVID-19 pandemic crisis has produced a “remarkable … mobilisation” of scientific knowledge, but at the same time has highlighted weaknesses in the design of global scientific cooperation and practices, despite lessons learned from previous health crises (Young, 2020). Knowledge sharing exists but there is concern about violation of open science principles, lack of reuse licences, low percentages of open data, wastage, duplication, misuse and retraction in COVID-19 research (Besançon et al., 2021; Glasziou, Sanders, and Hoffmann, 2020; Retraction Watch, n.d.)

To understand the ongoing impact of COVID-19 on scholarly communication, COKI’s tracking and analysis of institutional open research output enables us to identify patterns in research output, the location of researchers and the nature of research openness on a global scale. Through analysis of institutional demographic data, we explore the effects of COVID-19 on university workforce numbers and diversity of workforces. With the Open Knowledge Community coalition of stakeholders, we are growing, sustaining and maintaining the COKI data asset, developing a global community of practice around making change and sharing expertise on evaluation and openness.

References


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

The Royal Society (2021). Our journals are making all research findings and data relevant to the covid-19 pandemic immediately open access. https://royalsocietypublishing.org/topic/special-collections/novel-coronavirus-outbreak


Acknowledgements

The project is funded by the Research Office of Curtin University through a strategic grant, the Curtin University Faculty of Humanities, and the School of Media, Creative Arts and Social Inquiry.