Australian Research Data Commons

The Current and Future States of the eResearch Workforce

by the Australian Research Data Commons

17 July 2019

The Current and Future States of the eResearch Workforce

This document provides some background and explains the approaches taken to explore the supply and demand issues around digitally upskilling the Australian research workforce.

Data and software are fundamental to many areas of modern research practice. They are at the heart of eResearch - the utilisation of digital technologies[1] to support and conduct research. Yet ironically there is very little data available nationally and internationally about the volume, rate and complexity of research data and software being utilised or developed for research. This makes long term planning and capability development at a national scale challenging.

Taking the economic concept of "production" as a starting point, research activities typically involve a set of inputs and outputs. The inputs of research may include financial and human capital, instruments, raw materials for experiments, past research findings and datasets amongst other things. On the other hand the outputs of research are typically scholarly publications, but may also include data, software, computer models, prototypes, exhibitions, live performances and so on. One way in which the future demand of eResearch could be estimated is by examining the volume and rate of research productivity, both in absolute and relative terms.

In Information Sheet 1: Just how much data-related research is occurring: an analysis of Scopus database,[2] preliminary results from 2018 OECD International Survey of Scientific Authors (ISSA2) found that globally, 63% of sampled authors reported that new data and/or code were developed through research associated with their publications. The figure for Australian authors is 62%. This suggests that, in relative terms, the amount of new data and/or code generated during research (and thus eResearch activities) in Australia and globally is significant.

However, the OECD figure only provides a snapshot, point-in-time, view of the state of research data and/or code production. Using "data" as a search term for title, abstract and keywords of peer-reviewed articles, reviews and conference papers from Scopus database, bibliometric information for "data-related" scholarly publications is used to extrapolate the proportion of data-related research being undertaken globally and nationally. This serves as a proxy to gauge the volume and rate of eResearch activities over time. Over a period of thirty years, the growth of data-related publications, expressed as a percentage of total publications, has risen from 11.7% to 22.1% globally. In Australia, the growth is considerable from 13.4% in 1989 to 28.0% in 2018[3] — surpassing other R&D intensive countries such as the US, Canada, China and Singapore. Taking this trend as a baseline, it suggests that growth in data-related publications, and thereby eResearch activities, will likely continue in the near future.

The growth rate of research outputs, however, is not independent on the growth in the inputs, financial and human capital in particular. In Information Sheet 2: size of research support professional workforce in the Australian research sector, official statistics from the Australian Bureau of Statistics are examined to provide the relevant context to interpret the productivity growth. One clear explanation is that the size of the Australian

[1] An alternative characterisation uses "information and communication technologies" (see the Council of Australasian Directors of IT description).
[2] All Information Sheets were commissioned by the Australian Research Data Commons and written by Natalie Mast, N. Mast Consulting.
[3] Equivalently, the rate of change is, on average, 33% increase in every decade.

research community has increased considerably over time and therefore more scholarly publications (and data-related publications) were produced. Using "per researcher" to normalise the data, the cumulative (percentage) growth rate of data-related publications is demonstrated to be rising at a higher rate then the total publication rate.

One important observation from Information Sheet 2 is that there are considerable discrepancies across sectors in the "researcher to support-staff" ratio — in 2016, higher education is 1 to 0.4, government and private non-profit is 1 to 0.8, and business is 1 to 1.1. These statistics suggest that researchers in business and industry are better supported then those working in government, private non profit organisation, or high education. This raises some concerns whether the higher education sector is capable of retaining talents in eResearch support.

Finally in Information Sheet 3, the issue about the future demand of eResearch is placed in a broader context of the demand for data scientists and data related occupations across the domestic economy — specifically those professionals falling within "Computer System Design and Related Services"[4] and in the "Professional, Scientific and Technical Services" classifications. According to models developed by the Department of Jobs and Small Business[5] the projected growth for "Computer System Design and Related Services" is at 15.6% and "Professional, Scientific and Technical Services" is at 8.6%, "Tertiary Education" is only expected to grow by 3.9%, which is less than the overall economy. This suggests that within Australian universities eResearch support may be facing considerable competition from other sectors for well qualified data and software professionals.

To round out the discussion here, the three Information Sheets provide a partial view of the current and future states of eResearch. They should be viewed in tandem as starting points of a more considered investigation. Some further suggested issues to be considered include:

The Complexity of Research: research activities involving data and software are diverse and varying in complexity, and can be very discipline specific. Some research projects are extremely complex involving many steps in data collection and processing, many different data sources, (tens of) thousands line-of-code, many software packages, many cpu hours on a high performance computer, etc. The technical skills required to support a highly complex research project are not equivalent to skills required to support projects with a low level of data and software complexity. Indeed the complexity of research can be discipline specific and thus is driving some disciplines to introduce special data management (or scientific programming) courses as a part of their core postgraduate training.[6] This suggests that different disciplines may have different approaches to tackle the challenge of research complexity unique to their disciplines, and thus would have different requirements and expectations of the kind of eResearch support needed.

Research Quality and Impact: eResearch support is not merely about productivity in terms of sheer volume or rate of production. Key issues around research quality and impact (social, economic, cultural, environmental and health benefits beyond academia) are also important drivers of research, and thus eResearch. Low quality research data or software not only undercuts the quality of research, in extreme cases they can seriously undermine the validity of the research finding or render the research irreproducible.[7] In terms of research impact, the development and access to quality research data and software too can play an important role in the

[4] See ABS 1292.0 - Australian and New Zealand Standard Industrial Classification (ANZSIC), 2006 (Revision 2.0), Class 7000 Computer System Design and Related Services.
[5] See the Employment Outlook to May 2023 and the Labour Market Information Portal website. Note that the Department of Job and Small Business has been renamed as the "Department of Employment, Skills, Small and Family Business".
[6] See "Linguistic Data Management" by Nicholas Thieberger and Andrea L. Berez. *The Oxford Handbook of Linguistic Fieldwork*, edited by Nicholas Thieberger, 2011. DOI: 10.1093/oxfordhb/9780199571888.013.0005
[7] The recently released, *Management of Data and Information in Research: A guide supporting the Australian Code for the Responsible Conduct of Research*, by the ARC, NHMRC and Universities Australia, reiterates the essential role of good data stewardship in maintaining the integrity of research.

delivery of research impact.[8] High quality eResearch support and service are fundamental enablers of high quality and high impact research.

This document has examined some useful approaches based on the available data and statistics but it is clear that more work would be required to strengthen understanding of the issues involved.

[8] For a couple of Australian impact case studies where research data and algorithm played a pivotal role in the delivery of impact beyond academia, see the ARC Impact Case Studies portal:

- How the Joint Remote Sensing Research Program has improved land management in Australia
- Conserving and Interpreting Australia's Convict Past

All Information Sheets are commissioned by the Australian Research Data Commons and written by Natalie Mast, N. Mast Consulting

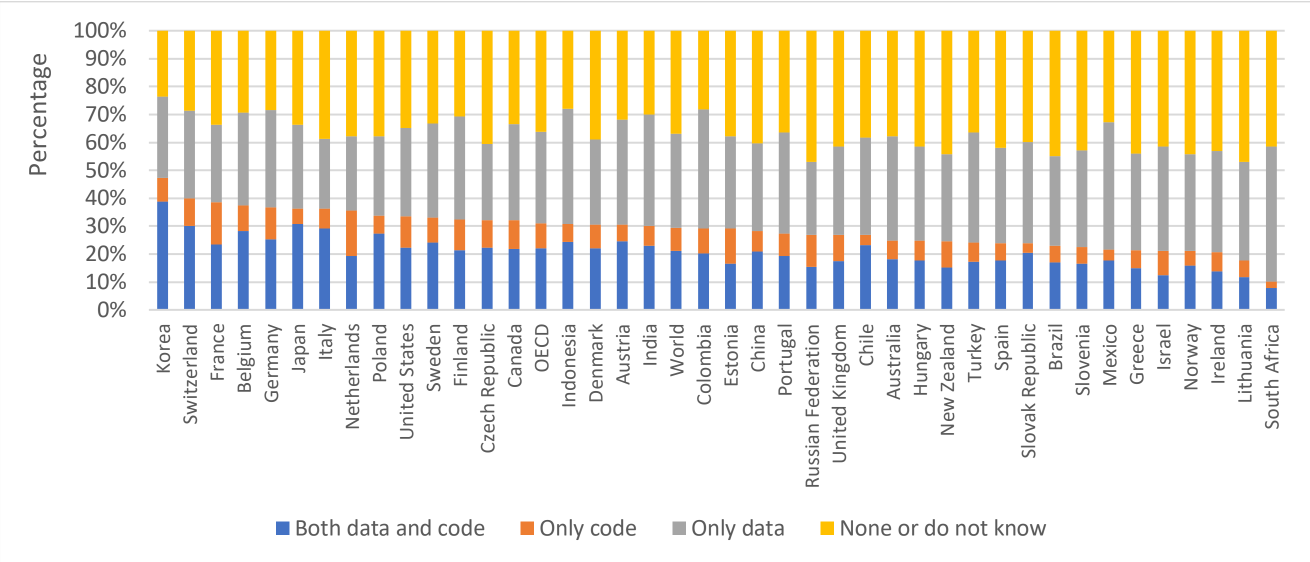
Information Sheet 1: Just how much data-related research is occurring – an analysis of Scopus documents

Research occurs in universities, governmental organisations, research institutes and within industry and the private sector. The manner in which the research produced is communicated and utilised differs across sectors. Peer-reviewed publication is the most common form of knowledge communication for researchers working within universities and research institutes.

Conducted by the OECD, the International Survey of Scientific Authors (ISSA) targets authors who have published in scientific peer-reviewed journals indexed by bibliometric provider, Scopus [1]. The survey was piloted in 2015 (ISSA1) (OECD, 2015), and also conducted in 2018 (ISSA2) (OECD, 2018), with only preliminary results from the second survey available at the time of writing. ISSA surveys are sent to a "randomly selected group of corresponding authors of scholarly documents" (OECD, 2019). Appendix 1 provides a summary of the questions from each survey considered relevant.

The preliminary results of ISSA2 found that globally, 63% of scientific production surveyed was expected to result in new data and/or code (see Figure 1), with the figure for the Australian scientific production surveyed being 62%.

Figure 1 Scientific production resulting in new data or code[2], by country of residence, 2017



Data Source: OECD ISSA2 https://doi.org/10.1787/888933930250

---

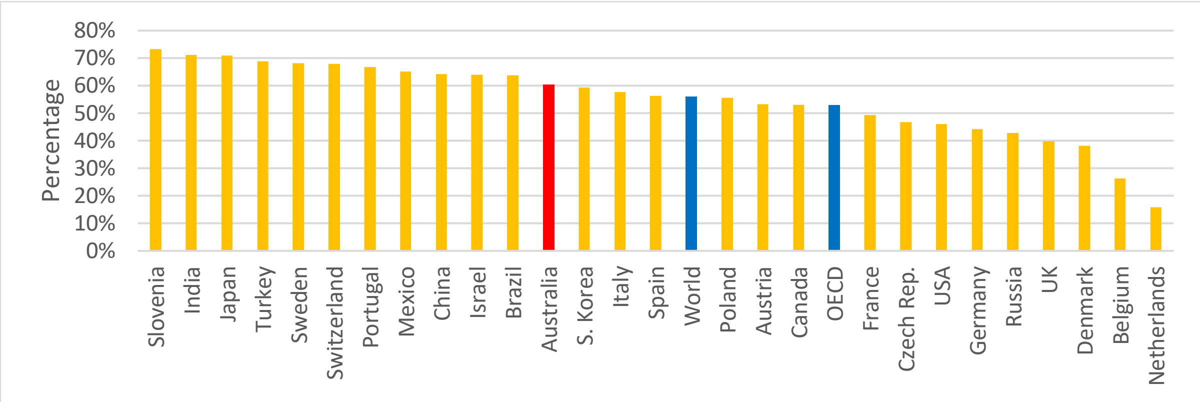
[1] Not all research fields have the same communication methods, for example publication by book and book chapter is common in the humanities and social sciences, while communication via conference proceedings is common in computer science and engineering. As a result, not all fields of research receive the same coverage by bibliometric providers such as Scopus, and thus results may be skewed towards the sciences.

[2] "Research data include numerical scores, textual records, images and sounds used as primary sources for scientific research. Codes include custom developed software and code, laboratory notebooks, and other computer enabled documents describing every step of the research work and protocols followed. Please exclude drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues

The OECD note that “the ISSA 2018 results show that the digitalisation of science is not limited to scientific fields that specialise in computer science or IT engineering. They also indicate significant potential for greater IT adoption in general scholarly practice as well as for harnessing the potential of data-driven research” (OECD, 2019, p. 155).

Using data from ISSA1 (Q17) the percentage of researchers who found databases and records to be important to their research can be calculated. Figure 2 shows that 60% of Australian researchers surveyed considered databases and records were important to their research.

Figure 2 Percentage of researchers[3] responding “Important” when questioned as to the importance of databases and records to their research



Data

Source: OECD ISSA1 https://survey2018.oecd.org/Survey.aspx?s=0c0dcc5298394efe9b202e4cd2296aa3

Bibliometric data relating to peer-reviewed publications can be used to extrapolate the proportion of data-related research being undertaken by researchers in research institutions and serve as a proxy for the level of eResearch currently being conducted. The next section focuses on Australia’s scholarly research output, using peer-reviewed articles, reviews and conference papers data from Scopus over the thirty-year period 1989-2018. Australia’s output[4] is compared to a global figure and benchmarked to China, USA, UK, Netherlands, Canada and Singapore.

Figure 3 shows that since 1989, Australia’s annual publication rate has undergone significant growth (6.8% p/a), at a rate 2.6% higher than that of the global average (see Appendix 2 for details of data extraction). Average growth rates in the last five years are lower for all countries examined as well as globally. Among the benchmark countries examined, Australia’s 3.3% growth rate is second only to China at 5.4% and well above the world average of 1.9%.

To determine the level of data-related research occurring, Scopus indexed publications were limited to those with the term “data” in the title, abstract or keywords. Over the last 30 years Australia has had an average growth rate of 9.6% in terms of data-related publications, which is 3% greater than the world

---

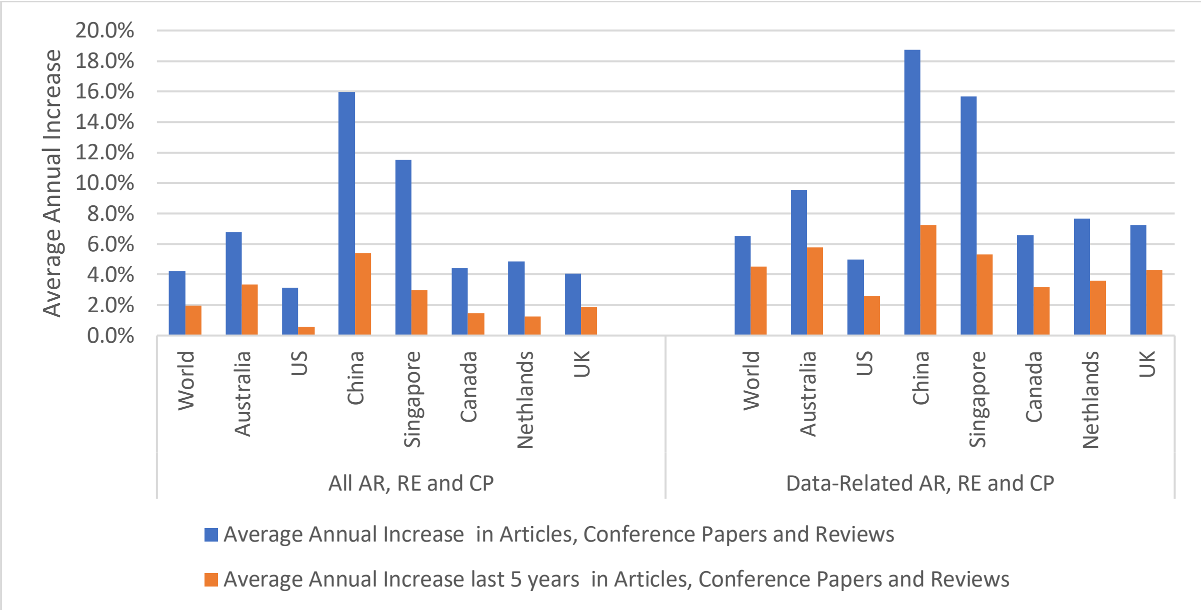
as well as physical non-digitised objects (e.g. laboratory samples, strains of bacteria and test animals such as mice, specimens, and archaeological findings)” (OECD, 2018, question 8).

[3] Only countries with at least 15 responses are shown. World and OECD totals were calculated for this report.

[4] To be classified as “Australian” a publication must have at least one institutional affiliation with an Australian address. The same rule applies to all other benchmarked countries. Publications involving authors from multiple countries are counted against each affiliated country listed.

average. If the time period examined is shorted to the last five years, Australia has an average growth rate of 5.8% in terms of data-related publications, which is 1.3% greater than the world average.

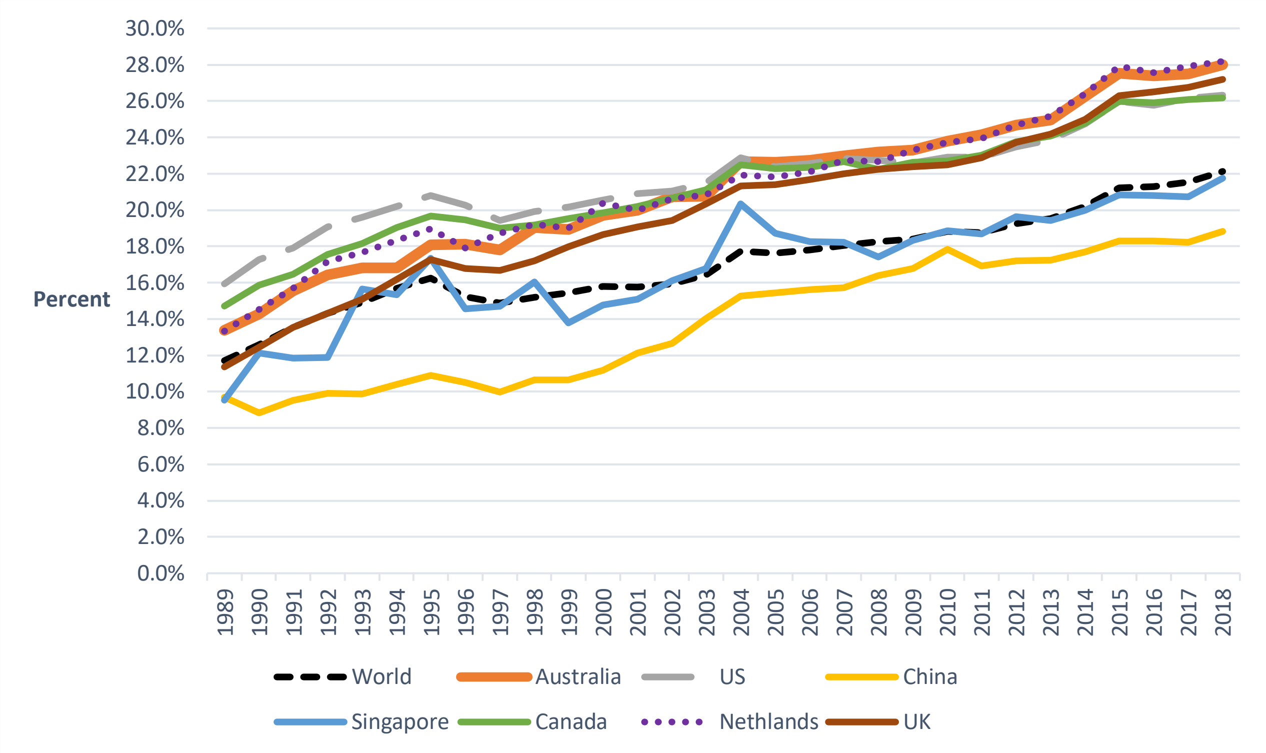
Figure 3 Average annual increase in Scopus indexed articles, reviews and conference papers, 1989-18



Data Source: Scopus, data extracted May 2019

Over the last thirty years the global proportion of the indexed output examined that can be considered data-related has almost doubled, with data-related publications growing from 11.7% to 22.1% of indexed articles, conference papers and reviews. The increase in the proportion of Australian publications deemed data-related has outpaced the global rate of increase, more than doubling from 13.4% in 1989 to 28.0% of the 2018 indexed output examined (see Figure 4).

Figure 4 Proportion of Scopus indexed articles, reviews and conference papers relating to data, 1989-18



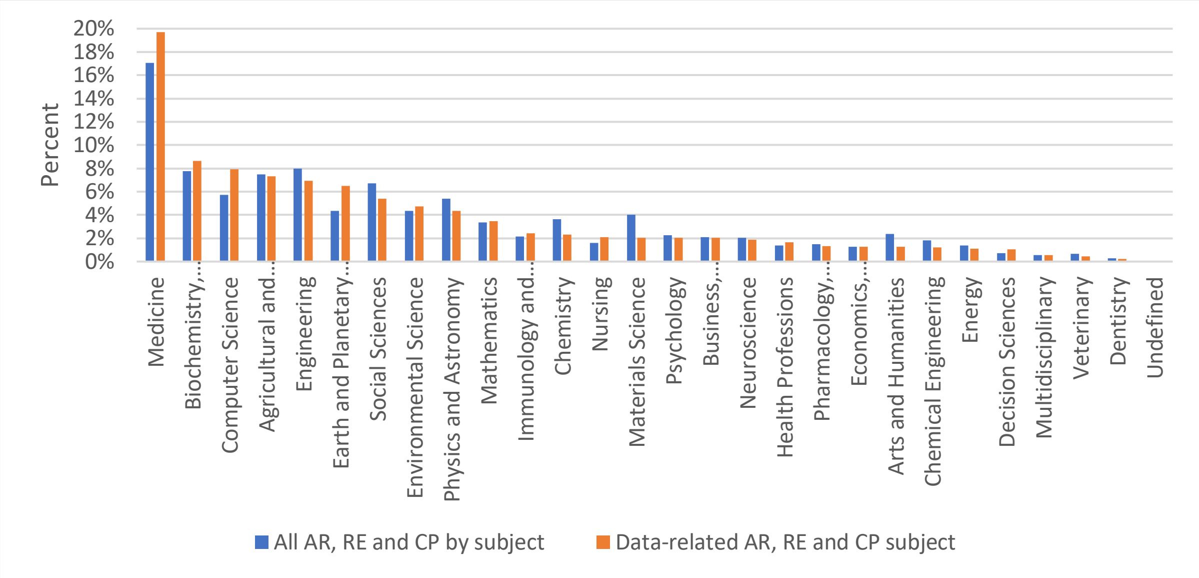
Data Source: Scopus, data extracted May 2019

Based on 2018 publications, Australian researchers were involved in 4.2% of the global data-related output examined, compared to 3.3% of all the Scopus indexed articles, reviews and conference papers from the same year (not shown).

Although not shown, it should be noted that the trend of an increasing proportion of data-related publications is very similar when examined through the filter of those publications with at least one Australian university by-line and those without a university by-line, suggesting that data-related research is on the increase across all segments of the Australian research sector that engage in peer-reviewed publication.

The Scopus data also shows which subject areas Australian data-related publications, produced over the last thirty years, are classified (see Figure 5). Scopus classify publications into 27 subject areas. It should be noted that publications can be classified in multiple subject areas and can therefore be counted more than once. Within Figure 5, the top 10 subject areas, from Medicine through to Mathematics, account for 75% of the data-related Australian research indexed by Scopus. Interestingly, if the time period is reduced to the last ten years (not shown below), while the relative size between the top ten subject areas for data-related publications changes, the subject areas remain the same and combined are also responsible for 75% of the data-related output. The same pattern is evident when the time period is reduced to the last five years.

Figure 5 Scopus subject areas for all publications and data-related publications 1989-2018



Data Source: Scopus, data extracted May 2019

In summary, over the last thirty years, the number of Scopus indexed articles, reviews and conference papers with at least one Australian based researcher has been increasing by an annual rate of 6.8%. The proportion of those publications which can be classified as data-related has increased from 13.4% in 1989 to 28% and in 2018, and Australia's proportion of data-related publications exceeded the global average by 5.9%.

Appendix 1: Key ISSA1 and ISSA2 Questions

ISSA Number	Question Number	Question	Type of Answer	data available
ISSA1	Q4	What is your current country of usual residence?	Alphabetical list of countries	Yes
ISSA1	Q7	In which field did you attain this qualification	searchable list of fields	yes
ISSA1	Q17	Could you please rate how important were the following sources for the research work that led to your article?	8 sub questions, binary response	yes
ISSA1	Q52	For the article under examination, has another researcher or organisation asked you for access to data or other research materials?	Y/N	yes
ISSA1	Q53	How have you responded to requests to share data, records or protocols related to your article?	Single choice with list of 6 options	yes
ISSA1	Q54	How would you respond to a request to share data or records related to your article?	Single choice with list of 6 options	yes
ISSA1	Q55	In your opinion, would another research team be able to attempt to replicate your research results with the information publicly available?	Single choice with list of 3 options and ability to provide comments	yes
ISSA2	Q4	*Where is your usual place of residence?	Alphabetical list of countries	no
ISSA2	Q8	Did you or your co-authors develop new data or computer enabled code (?) as part of the work for this paper?	single choice list of 4 options	yes
ISSA2	Q9	Have you or your co-authors deposited them or made available in a generic or specialised repository? (e.g. a library, archive, online platform)	single choice list of 4 options	no
ISSA2	Q10	Have you or your co-authors delivered them to a journal or publisher as additional material to support publication?	single choice list of 4 options	no
ISSA2	Q11	Have you or your co-authors made them directly available to fellow researchers?	single choice list of 4 options	no
ISSA2	Q12	Which of the following apply to data and code coming out of your research work for this paper?	multiple choice list of 7 options	no
ISSA2	Q13	Please indicate whether the following factors have significantly constrained or enhanced the level of access granted to the research outputs from this paper (i.e. data, codes and/or publication).	8-part question with single choice of 3 options	no
ISSA2	Q14	Over the last twelve months, did you make use of any online platforms or related apps, tools or solutions for any of the following as part of your scientific/research work?	4-part question with multiple choice in each list of options	no
ISSA2	Q15	Which of the following activities best describe the scientific/research work that you have been involved in over the last twelve months?	multiple choice list of 6 options	no
ISSA2	Q16	As part of your core scientific/research activities, have you or your team been dealing with any of the following over the last twelve months.	6-part question, single choice of 4 options	no
ISSA2	Q17	Over the last twelve months, have you worked with data...?	multiple choice list of 9 options	no
ISSA2	Q18	Which of the following has proved to be most challenging in your scientific/research work?	single choice list of 4 options	no
ISSA2	Q19	Which of the following skills have you found most important for your scientific/research work?	option to select 2 items from a list of 6 options	no
ISSA2	Q20	Which of the following information and communication infrastructures have you found most important for your scientific/research work?	option to select 2 items from a list of 8 options	no
ISSA2	Q30	which was the main subject of this qualification	List of six disciplines with subdisciplines beneath	no
ISSA2	Q26	(what are your views on) The trend towards an increasing use of digital tools in science and research…	respondents split into two groups. 6-part question with each question having a left/right statement and a 10-point scale separating them.	no

Data Source: OECD ISSA1 https://survey2018.oecd.org/Survey.aspx?s=0c0dcc5298394efe9b202e4cd2296aa3; OECD ISSA2 https://doi.org/10.1787/888933930250

Appendix 2: Scopus Search

Using the search option on the Scopus search page:

Primary search function to determine global figures *=PUBYEAR AFT 1988 AND PUBYEAR BEF 2019 AND DOCTYPE (AR) OR DOCTYPE (RE) OR DOCTYPE (CP)*

To limit to individual countries *=PUBYEAR AFT 1988 AND PUBYEAR BEF 2019 AND DOCTYPE (AR) OR DOCTYPE (RE) OR DOCTYPE (CP) AND AFFILCOUNTRY (insert name of country)*

Primary search function to determine global figures for data focused research *=TITLE-ABS-KEY (data) AND PUBYEAR AFT 1988 AND PUBYEAR BEF 2019 AND DOCTYPE (AR) OR DOCTYPE (RE) OR DOCTYPE (CP)*

To limit to individual countries *=TITLE-ABS-KEY (data) AND PUBYEAR AFT 1988 AND PUBYEAR BEF 2019 AND DOCTYPE (AR) OR DOCTYPE (RE) OR DOCTYPE (CP) AND AFFILCOUNTRY (insert name of country)*

NOTE: Searches of bibliometric databases are a snapshot of the database at a particular point in time. Bibliometric providers are continually indexing and adding to their database, not just in terms of updating citation metrics, but in expanding the coverage of journals and other publication sources. Thus, two identical extracts run weeks apart will generally yield only similar rather than exact results.

References

Elsevier, Scopus, (subscription only) https://www.scopus.com/home.uri

OECD, 2015, *International Survey of Scientific Authors (ISSA*), Available from: https://www.oecd.org/science/survey-of-scientific-authors.htm

OECD, 2018, *OECD International Survey of Scientific Authors 2018, "How is science going digital?" Available from:* https://www.oecd.org/science/survey-of-scientific-authors.htm

OECD, 2019, *Measuring the Digital Transformation: A Roadmap for the Future,* OECD Publishing Paris.

Information Sheet 2: Size of research support professional workforce in the Australian Research Sector

There are two ways in which research output, in terms of indexed publications[1], increase on an annual basis. The first way is via an increase in the size of the research community actively engaged in publication, while the second method is via an increase in productivity within the research community, whereby a greater number of publications per researcher is produced annually.

The Australian Bureau of Statistics (ABS) Higher Education Expenditure on R&D (HERD) collection (ABS, 2018a), which has data available from 1992 onwards, can be utilised to determine the size of the higher education research sector in Australia. The HERD collection provides data on academic staff, support staff (technicians, skilled and unskilled employees, secretarial and clerical staff), and postgraduate students directly involved in research and development activity[2]. The ABS also undertake similar exercises for the government and private non-profit organisations (G/PNFP)[3], as well as the business sector (ABS, 2018b and ABS, 2018c)[4]. Data is provided as person years of effort (PYE)[5]. 2016 is the latest year of data available for all three sectors, with the next release, measuring 2018 PYE, expected in 2020.

Figure 1 shows that among the three sectors examined, changes in the size of the workforce has not been consistent. Focusing on the period 2010-2016, there has been a small decrease in the size of the researcher cohort working within the G/PNFP sector (-2%), while the support staff cohort working in the G/PNFP sector has declined at a greater rate (-8%). There has been substantial growth in both cohorts operating within the business sector, with the size of the researcher cohort increasing by 18%

---

[1] Excluding the possibility of a bibliometric provider expanding the source list indexed.

[2]An examination of the Department of Education's annual *Higher Education Staff Data Collection* was also undertaken. However, this collection shows that in recent years there has been a significant decline in the number of staff members classified within the non-academic "research other" classification, which has traditionally been used to classify research support staff. Rather than a real decline in research support staff (which has not been evident in the ABS' HERD collection), it appears that many research support staff are now being classified within the non-academic "other function", which has seen a significant increase in FTE that more than covers the decline in non-academic "research other" numbers. At the same time, there has also been a significant increase the number of academic Level A research only staff, which may include individuals formerly classified within the "research other" category being transferred into academic research only positions (Department of Education 2018).

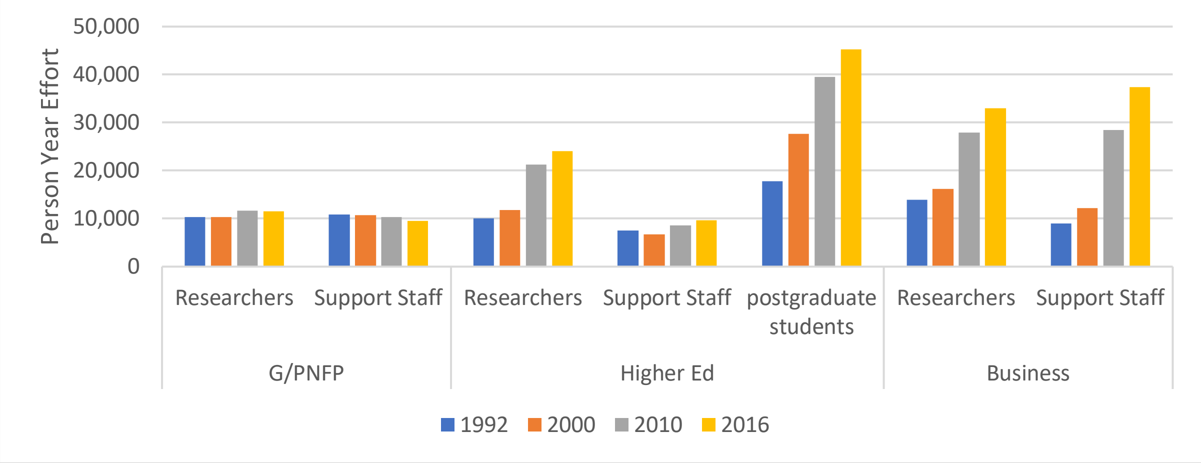
[4]Note: 8104.0 Research and Experimental Development, Businesses, Australia was collected annually until 2011-12.

[5] Since 1994 the ABS has not split "technical staff" and "other staff" in the HERD collection, and thus this analysis merges those two groups into "support staff" when looking at the government and private not-for-profit as well as the business collections to allow for comparison. The ABS define "other staff" as: Technicians, skilled and unskilled employees, secretarial and clerical staff directly associated with R&D activity.

Within this report, the HERD classification "other staff" is equated with research "support staff". Likewise, the HERD uses the term "academic staff", but the two other ABS collections refer to "researchers". To avoid confusion the term researchers is used in relation to the higher education sector in this report.

since 2010, and the support staff cohort increasing by 31%. Within the higher education sector, the researcher cohort (13%) as well as the support staff cohort (14%) have grown by a similar rate.

Figure 1 Australian researchers, support staff and HDR student PYE per sector, 1992, 2000, 2010 and 2016



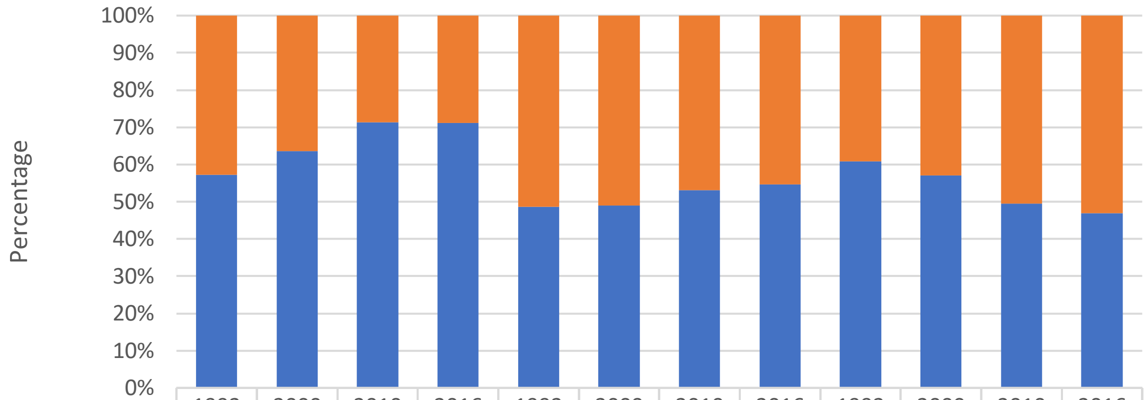
Data Source: Australian Bureau of Statistics (ABS), 2018, 8111.0; 8109.0; 8104.0 and past releases

Within Figure 1, both the number of researchers and support staff in the G/PNFP sector have decreased since 2010. The ABS break down the data for the G/PNFP sector into subsectors (not shown here). The private not-for-profit data, which focuses on organisations such as medical research institutes, showed an increase in researcher PYE of 24% and an increase in support staff of 8% over the period 2010-16. The decline within the G/PNFP sector is being driven by reduced staff numbers in the government sector at both the state and commonwealth levels. Research PYE fell by 14% at the Commonwealth level and 7% at the state/territory level. Support staff levels fell by 23% at the Commonwealth level although there was an increase of 3% at the state/territory level.

While the growth in researchers working in the higher education sector (13%) and the business sector (18%) are similar, the same cannot be said for support staff levels, which increased by 14% in the higher education system, compared to 31% in the business sector. One reason for this difference may be the significant growth in the higher degree by research (HDR) population in the higher education sector, which also saw an overall increase of 14% between 2010 and 2016. HDR students may be undertaking work traditionally carried out by support staff and thus have somewhat mitigated the employment of support staff in the higher education sector.

Since 1992 the ratio of support staff to researchers has decreased in both the higher education and G/PNFP sectors, particularly in the higher education sector, but the ratio has increased considerably in the business sector (see Figure 2). Figure 2 shows that between 2000 and 2010, the ratio of support staff to researchers within the Higher Education sector decreased but was steady when 2010 is compared to 2016. The business sector has seen the ratio of support staff to researchers increase in each of the four snapshots examined. Meanwhile the G/NFP sector ratio of support staff to researchers was close to steady between 1992 and 2000, decreased in 2010 and was again relatively steady in 2016. Over time, the three segments of the Australian research sector examined have not behaved uniformly in terms of recruitment of both researchers and research support staff.

Figure 2 Proportion of Australian researchers to support staff per sector, 1992, 2000, 2010 and 2016



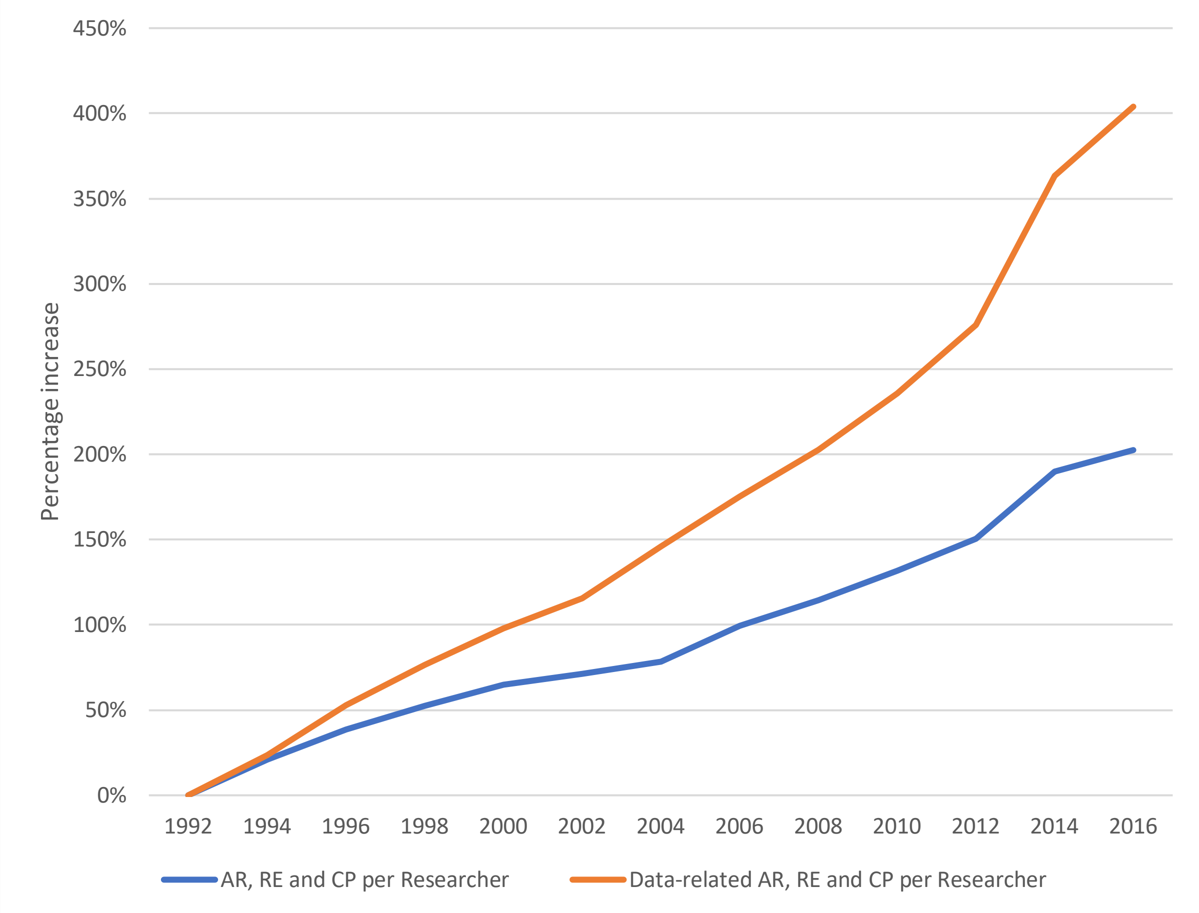
	Higher Ed				G/PNFP				Business			
	1992	2000	2010	2016	1992	2000	2010	2016	1992	2000	2010	2016
Support Staff	7,504	6,780	8,530	9,720	10,872	10,669	10,271	9,488	8,976	12,170	28,521	37,451
Researchers	10,059	11,854	21,300	24,075	10,301	10,273	11,675	11,463	13,943	16,221	27,990	33,016

Data Source: Data Source: Australian Bureau of Statistics (ABS), 2018, 8111.0; 8109.0; 8104.0 and past releases

Just as recruitment processes differ across the sub-sectors of the Australian research sector, so too do the outcomes of research. While the business sector research community is large, in fact it is larger than the higher education (excluding HDR students) and G/NFP sectors combined, the business sector does not generally communicate its research via indexed peer-reviewed publications. Including business sector researchers in a measure of productivity based upon scholarly publication methods significantly increases the denominator with a negligible increase to the numerator. Likewise, as research support staff are not normally generating their own research and are not necessarily included in the author list of a publication, their inclusion in a productivity measure increases the denominator, without a corresponding increase in the numerator. The HDR cohort is the largest group in the ABS data, yet not all research students publish during their candidature. Further, HDR authored output cannot be identified using the data provided by Scopus. Thus, researchers within the Australian business sector, research support staff from all three sectors examined, and HDR students will be excluded from the next part of this analysis.

Figure 3 charts the rate of change in Scopus indexed articles, reviews and conference papers per Australian researcher (working within the higher education and G/PNFP sectors) over time, and also provides figures on the change in data-related publications per researcher. As expected, given the higher rate of growth for data-related publications compared to all Australian indexed output (see Information Sheet 1 Figures 3 and 4), the increase of data-related publications per researcher outpaces that of all indexed output. Figure 3 shows that by 2016 the increase since 1992 in the proportion of data-related publications per researcher was twice that of the increase in all indexed publications per researcher.

Figure 3 Increase in Scopus indexed publications and data-related publications by Australian researcher (higher education & G/PNFP sectors) over time, 1992-2016



Australian Bureau of Statistics (ABS), 2018, 8111.0; 8109.0; and past releases and Scopus, data extracted May 2019.

Although publication output is considered an indicator of productivity for researchers, and not all research support staff are involved in the publication process, publications per support staff are also increasing. Given the overall growth rate for support staff in the combined higher education G/PNFP sectors (see Figures 1 and 2) has not matched the level of growth in researchers, the publications per support staff growth rate exceeds the increases for researchers seen in Figure 3.

In summary, Australian researchers in the higher education and G/PNFP sectors are producing more indexed items each year. This increase appears to be driven by an increase in researchers in the higher education sector, as well as an increase in indexed output, with a greater number of refereed articles/reviews/conference papers being produced per researcher. Driven by the increasing proportion of indexed items that are data-related, the rate of increase in data-related publications per researcher has outpaced that of all items per researcher by a factor of two.

References

Australian Bureau of Statistics (ABS), 2018a, *8111.0 - Research and Experimental Development, Higher Education Organisations, Australia, 2016 (and past releases)* https://www.abs.gov.au/AUSSTATS/abs@.nsf/second+level+view?ReadForm&prodno=8111.0&viewtitle=Research%20and%20Experimental%20Development,%20Higher%20Education%20Organisations,%20Australia~2016~Latest~22/05/2018&&tabname=Past%20Future%20Issues&prodno=8111.0&issue=2016&num=&view=&

Australian Bureau of Statistics (ABS), 2018b, *8109.0 - Research and Experimental Development, Government and Private Non-Profit Organisations, Australia, 2016-17 (and past releases)* https://www.abs.gov.au/AUSSTATS/abs@.nsf/second+level+view?ReadForm&prodno=8109.0&viewtitle=Research%20and%20Experimental%20Development,%20Government%20and%20Private%20Non-Profit%20Organisations,%20Australia~2016-17~Latest~05/07/2018&&tabname=Past%20Future%20Issues&prodno=8109.0&issue=2016-17&num=&view=&

Australian Bureau of Statistics (ABS), 2018c, *8104.0 - Research and Experimental Development, Businesses, Australia, 2015-16 (and past releases)* https://www.abs.gov.au/AUSSTATS/abs@.nsf/second+level+view?ReadForm&prodno=8104.0&viewtitle=Research%20and%20Experimental%20Development,%20Businesses,%20Australia~2015-16~Latest~15/09/2017&&tabname=Past%20Future%20Issues&prodno=8104.0&issue=2015-16&num=&view=&

Department of Education, 2018, *the Higher Education staff data collection* https://www.education.gov.au/staff-data.

Elsevier, Scopus, 2019, (subscription only) https://www.scopus.com/home.uri

Information Sheet 3: Anticipated future growth in demand for workers with eResearch support skills

Determining the exact number of people involved in eResearch support, that is staff supporting eResearch including data stewards and research software engineers, as well as how many eResearch support staff will be required in the future is not a simple endeavour. This is because many of those involved in eResearch support working within the higher education, government and private not-for-profit (G/PNFP) research sectors are not easily identified (see Information Sheet 2). However, Commonwealth data and a number of studies focused on the expansion of data science occupations and the information and communications technology (ICT) sector can be used to begin to establish a picture of the current state of play and future requirements.

In 2018, the Commonwealth Department of Jobs and Small Business[1] released a report on the *Employment Outlook to May 2023*. The report looks at both industries and occupations and found that: “total employment is projected to increase by 886,100 (7.1 per cent or 1.4 per cent per annum) over the five years to May 2023” (Department of Jobs and Small Business, 2018, p.3).

e-Research support staff will predominately be located in “Computer System Design and Related Services” and in the “Professional, Scientific and Technical Services”. The Department of Jobs and Small Businesses note in regard to “Professional, Scientific and Technical Services”:

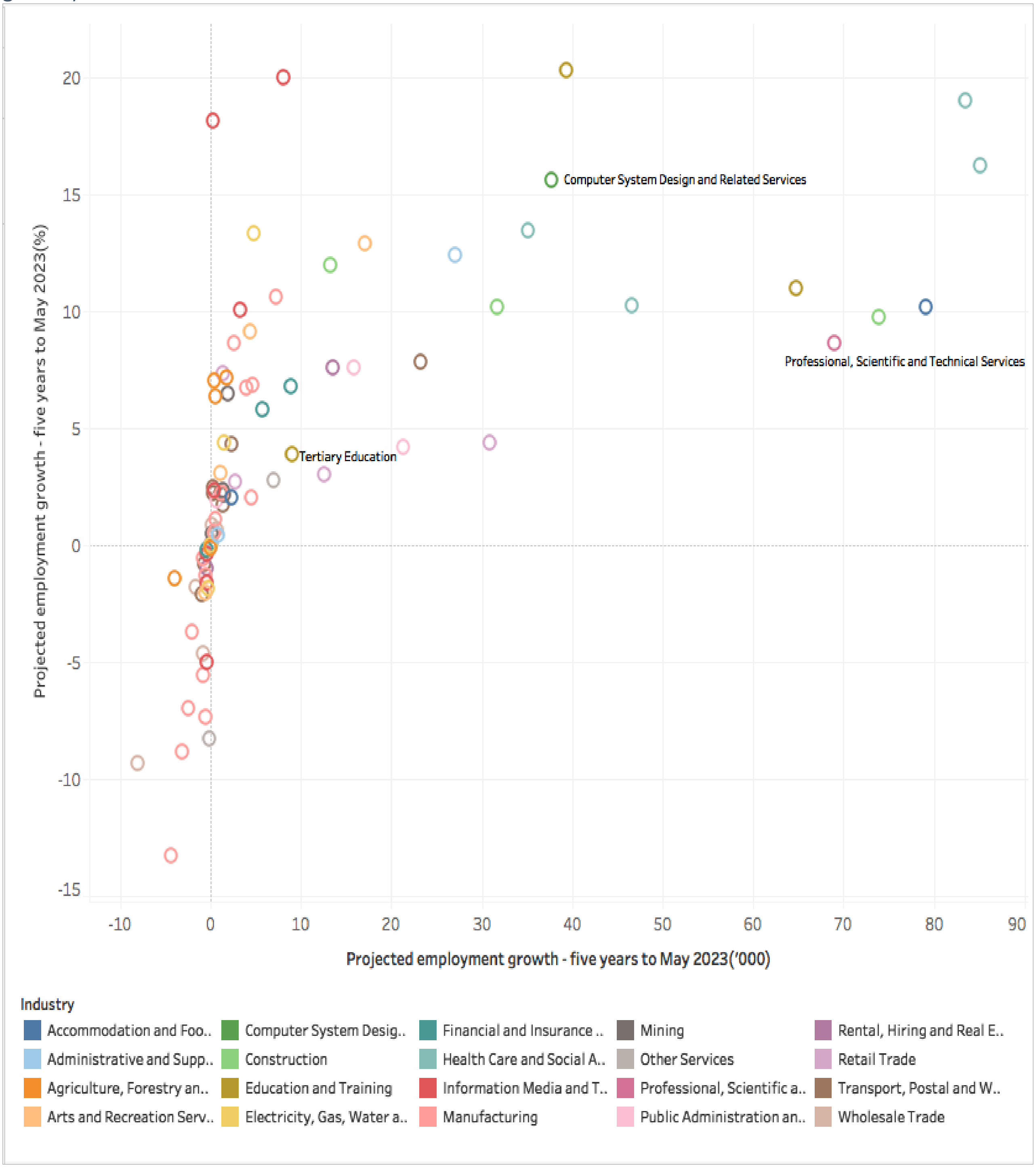
> This is a large industry, with over one million workers…. Many of (the services provided) support other industries. The workforce is highly skilled, and workers usually have a university degree (Department of Employment, Skills, Small and Family Business, 2019).

The data is classified using Australian and New Zealand Standard Industrial Classification (ANZSIC) codes, which provides data at four levels: division; subdivision; group; and, class. Figure 1 shows the projected employment changes to May 2023 by subdivision, which are coloured according to the division in which they classified.

The data shows that “Computer System Design and Related Services” are expected to grow by 15.6% (more than twice the overall rate) over the five-year period, with “Professional, Scientific and Technical Services” projected to grow by 8.6%. The “Tertiary Education” industry, which includes academics on teaching and research contracts, is also expected to grow by 3.9%, which is less than the overall economy.

[1] As a result of administrative arrangements, the Commonwealth Department of Jobs and Small Business underwent a name change at the end of May 2019 and is now known as the Department of Employment, Skills, Small and Family Business.

Figure 1 Projected employment changes in Australia to May 2023 by industry (‘000s and percentage growth)



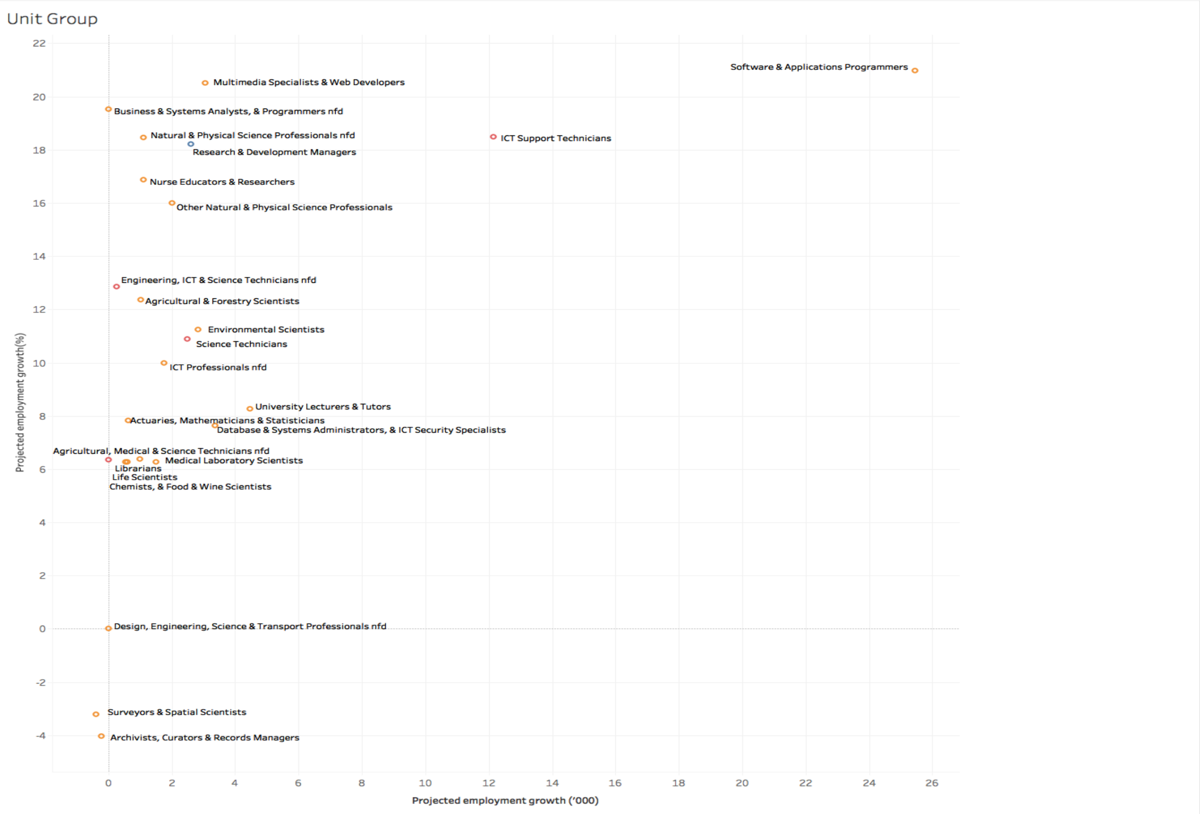
Data Source: Department of Jobs and Small Business http://lmip.gov.au/PortalFile.axd?FieldID=2787734&.xlsx

The employment outlook also examines projected growth in occupations, which can be included in multiple industries. For example, in the case of “Software and Application Programmers”, just over 60% are employed in the “Professional, Scientific and Technical Services industries” (Department of Jobs and Small Business, 2018, p.6).

Figure 2 shows the projected employment changes by occupation, limited to those likely to be involved in the production and support of research. The data is classified using the Australian and New Zealand Standard Classification of Occupations (ANZSCO), which employs five levels of classification: major group; sub-major group; minor group; unit group; and occupation. Figure 2 has been refined to only include those unit groups classified within the major fields of “Management”, “Professionals”, and “Technicians & Trades Workers”. The data was then further refined in an attempt to capture those involved in the research sector and includes occupations with the words researcher and scientists in their title, as well as those ICT and Business, Human Resource and Marketing Professionals related roles considered to be relevant to eResearch support. It should be noted that some of the definitions used in the ANZSCO, for example in relation to Librarians, tend to have a very traditional view of the occupation and are not a strong fit when it comes to identifying eResearch support staff, even though a proportion of staff working in the occupations are providing eResearch support services.

Within Figure 2, with the exception of “Surveyors and Spatial Scientists”, “Archivists, Curators and Record Managers” and “Design, Engineering, Science and Transport Professionals nfd[2]”, growth for all occupations is over 6% and in the majority of cases, over 10 percent, which significantly exceeds the projected 7.1% growth rate projected for the entire economy. While the percentage growth is almost universally strong among the occupations shown, the translation of that growth into workers is not uniform. Given the diverse number of industries which they serve, including the research sector, it is not surprising that the number of software and applications programmers is projected to grow by over 25,000. By way of comparison, Engineering, ICT and Science Technicians nfd, are anticipated to increase by around 13%, but this will result in an increase of less than 300 individuals.

Figure 2 Projected employment changes in Australia to May 2023 by occupation, (‘000s and percentage growth)



Data Source: Department of Jobs and Small Business http://lmip.gov.au/PortalFile.axd?FieldID=2787735&.xlsx

[2] nfd = not further defined.

In February 2018, Deloitte Access Economics (DAE), undertook a study entitled *The Future of Work: Occupational and Education Trends in Data Science in Australia*. DAE determined that the Australian data science workforce in the financial year 2016-17 was 301,000 people (DAE, 2018, p. 6). Using the Australian and New Zealand Standard Classification of Occupations (ANZSCO), the DAE figure is determined using data on occupations that are specifically related to key IT roles, and, a proportion of workers within key industries that require IT specialisation, such as Financial and Insurance Services; Agriculture, Forestry and Fishing; and, Professional, Scientific and Technical Services to provide a figure for the "Non-IT occupations grouping"[3].

The DAE report projected an annual growth rate of 2.4% over a five-year period to 2021-22, representing an increase of around 38,000 people in the data workforce (Economics 2018b). The DAE report also breaks down the areas of occupation in which those in data science employment are likely to be found.

Table 1 shows the Australian data science workforce by occupation. eResearch support staff are most likely to fall into the "Non-IT occupations grouping", which is anticipated to grow by an annual 3.2% over the next five years (DAE, 2018, p.7).

Table 1 Deloitte Access Economics' data science employment forecasts by occupation, 2016-17 to 2021-22

Occupation	2016-17 ('000)	2012-22 ('000)	Change in Employment ('000)	Average Annual Growth Rate
ICT Managers	59.6	65.9	6.3	2.0%
Actuaries Mathematicians and Statisticians	7.3	8	0.8	2.0%
ICT Business and Systems Analysts	29	30.2	1.2	0.8%
Software and Applications Programmers	108.4	125.9	17.5	3.0%
Database and Systems Administrators and ICT Security Specialists	45	50.5	5.5	2.3%
Computer Network Professionals	28.9	31.7	2.8	1.9%
Non-IT occupations grouping	22.7	26.6	3.8	3.2%
Total Data Science	**300.9**	**338.8**	**37.9**	**2.4%**

Data Source: DAE, The future of work: Occupational and education trends in data science in Australia, p.11

The DAE report allows for an analysis of the "Non-IT occupations grouping" by the occupations falling within the "Professional, Scientific and Technical Services" industry classification. The DAE report estimates that 18 percent of the "Professional, Scientific and Technical Services" industry is engaged in data science work (DAE, 2018, p.11) and that data science employment within the industry is expected to grow at 3.8% per annum (see Table 2). Within Table 2, a breakdown of occupations shows that a number, including "Life Scientists" (5.1%), "Environmental Scientists" (4.2%) and "Geologists and Geophysicists" (4.5%) are projected to grow at a significantly greater rate than that projected for the overall Data Science field by 2012-22 (2.4%) (DAE, 2018, p.11).

[3] "To account for this in our employment forecasts, we have used previous research (Burning Glass Technologies, 2017) which estimated the share of job openings that comprise data science and analytics roles in particular industries and applied these proportions to our workforce estimates for these non-IT occupations by industry" (DAE, 2018, p. 5).

Table 2 Breakdown of Deloitte Access Economics' data science employment forecasts by occupations classified within Total Professional, Scientific and Technical Services, 2016-17 to 2021-22

Occupation	2016-17 ('000)	2012-22 ('000)	Change in Employment ('000)	Average Annual Growth Rate
Economists	0.4	0.4	0.1	3.1%
Cartographers and surveyors	1.3	1.6	0.4	5.3%
Urban and Regional Planners	0.6	0.7	0.1	2.7%
Life Scientists	0.5	0.7	0.1	5.1%
Management and Organisation Analysts	4.4	5.1	0.8	3.3%
Environmental Scientists	1.5	1.7	0.3	4.2%
Geologists and Geophysicists	0.6	0.7	0.1	4.5%
Medical Laboratory Scientists	1	1.1	0.1	2.2%
Total Professional, Scientific and Technical Services	**10.3**	**12**	**2**	**3.80%**

Data Source: DAE, The future of work: Occupational and education trends in data science in Australia, p.11

In summary, although not specifically identified, demand for workers with many of the skills required for eResearch support roles is expected to grow over the next five years at a rate greater than that projected for the overall economy.

There are external factors that can also influence the level of demand for eResearch support. One factor which may slow down demand would be an economic downturn, which could limit the level of Government funding available for research, reduce industry investment in research and limit universities and institutes from investing in research projects, staff and equipment. Conversely, a prioritisation of data driven research through government initiatives and the redirection of funding could lead to the rate of increase in demand for eResearch support accelerating. Another issue that could significantly impede the rate of eResearch support growth is competition from other sectors employing skilled individuals,[4] leading to a situation of unmet demand.

In considering the future demand for workers with eResearch support skills, a number of factors should be taken into consideration, including:

- the growth rates in researchers, support staff and HDR students the ABS data highlighted for the higher education, business, and the government/PNFP sectors (see Information Sheet 2);
- the growth rates identified in data-related publications (see Information Sheet 1);
- the projections from the Commonwealth's employment outlook;
- DAE's data science employment forecasts; and,
- reports of an increasing demand for workers with skills sets similar to some of those required by eResearch support staff (see ft. 4)

As a result, taking into consideration the current size of the eResearch support workforce (N), modelling a low (0.02), medium (0.03) and high rate (0.04) of growth (X) over time (^{t}) would be prudent. Therefore, the projected future size of the eResearch support workforce (S) could be expressed as:

[4] See for example: Elsevier & Mori, 2019; Pompa & Burke, 2017; AIIA, 2017; Alpha Beta, 2019; Deloitte, 2019; and Crozier, 2018.

$$S=N(1+X)^t$$

The size of the eResearch support workforce is not quite the same as the demand level for the eResearch support workforce. There well might be unmet demand in the research sector where even with the application of the high growth rate outlined above, not all eResearch support requirements are being met, particularly if there is an unexpected surge in the proportion of data-related research requiring support.

References

Alpha Beta, 2019, Future Skills. Available at:www.alphabeta.com/wp-content/uploads/2019/01/google-skills-report.pdf

Australian Information Industry Association, 2017, *AIIA: Skills for Today. Jobs for Tomorrow. Available at:* https://aiia.com.au/__data/assets/pdf_file/0020/81074/SKILLS-FOR-TODAY-JOBS-FOR-TOMORROW-FINAL.pdf

Crozier, R., 2018, *Rio Tinto faces big shortfall of data scientists*. Available from: https://www.itnews.com.au/news/rio-tinto-faces-massive-shortfall-of-data-scientists-487404 .

Deloitte, 2019, The Path to Prosperity: Why the Future of Work is Human, Available at: http://images.content.deloitte.com.au/Web/DELOITTEAUSTRALIA/%7b90572c4f-4bb5-4b54-bf27-cc100d86890d%7d_20190612-btlc-inbound-future-work-human-report.pdf

Deloitte Access Economics, 2018, *The Future of Work: Occupational and Educational trends in Data Science Australia*. Available at: www2.deloitte.com/content/dam/Deloitte/au/Documents/Economics/deloitte-au-economics-future-of-work-occupational-education-trends-data-science-170418.pdf

Department of Jobs and Small Business*, 2018,* ***Employment Outlook to May 2023***. Available at: lmip.gov.au/default.aspx?LMIP/GainInsights/EmploymentProjections.

Department of Employment, Skills, Small and Family Business, 2019, Jobs Outlook Web Resource: https://joboutlook.gov.au/IndustrySpecific?search=Industry&Industry=M

Elsevier & Mori, 2019, *Research Futures: Drivers and scenarios for the next decade*. Available at: www.elsevier.com/connect/elsevier-research-futures-report

Pompa, C., & Burke, T., 2017, *Data Science and Analytics Skills Shortage: Equipping the APEC Workforce with the Competencies Demanded by Employers*. Available at: www.apec.org/Publications/2017/11/Data-Science-and-Analytics-Skills-Shortage