Surveying the scale of the research-IT support workforce.

A survey and report commissioned by the Australian Research Data Commons (ARDC)
July 2019

Dr Markus Buchhorn, eResearch consultant - m.buchhorn@outlook.com - 0417 281 429

Introduction

Nationally we do not have any strong sense of the scale of the workforce involved in supporting researchers' use of data, tools and platforms, nor of the needs of that workforce. Yet this workforce is fundamental to modern research techniques, and to Australia's global competitiveness in the research endeavour. Many institutions, organisations and funding programs invest in research support systems, including data, software, computing platforms and networks. The importance of the support workforce around those systems and in direct support of researchers is becoming increasingly clear.

In commissioning a survey of the workforce, the main priority for ARDC was to measure the scale of the community that particularly support researchers in their use of data and tools, in a variety of related roles. The intention of the survey is to support future discussions of the capacity and capabilities of the workforce, their skills needs, and their career paths. Following discussions with other stakeholders in the "eResearch" support community it was decided to further expand the survey into support for infrastructure and training.

This report outlines the approach taken to develop a meaningful survey of the workforce, the processes to communicate with the community, the analysis of the results from a very diverse community of organisations and institutes, and presents the results across a range of roles of working arrangements. Many challenges were identified throughout the survey and required some mitigation, or just recognition. It must be acknowledged up front that the measurements are underestimates of the totals, and classifications are blurred. At the same time, the survey provides a useful starting point for further analysis and perhaps future surveys, and it is the first broad-scale overview of the national research IT support workforce, and perhaps is the first of its type anywhere.

Developing the survey

There are many challenges in undertaking a survey such as this. The survey had to be meaningful and user-friendly, i.e. provide sufficient coverage across categories of organisations and roles, and attract sufficient participation, so that the statistics are sufficiently reflective of the current situation.

Boundaries

Essentially, we need to decide who do we want to count, and who do we not want to count. The survey should measure the workforce working directly with researchers, and not the much larger halo of ancillary support services. Even though the latter may still be crucial for some research, they also typically support a wider range of other users. Research support comes from a spectrum of providers, formally and informally. They include project-level support, from both support staff (formally) and researchers and students (informally). They include various tiers of institutional support through e.g. functional areas such as IT departments, libraries and various eResearch groups, and extend up through state-agency support and ultimately a range of national and even international initiatives. Those service providers are themselves further supported by additional

technical and administrative functions. At the same time, support-providers may also support other functions within the sector, e.g. education and administration in a university context.

There are edge-cases. One example is identity provision, which is fundamental to collaboration and data sharing, yet the management of identity/attributes is very broadly applicable beyond research. Then again particular authorisation and access mechanisms that depend on it may be very specific to research activities. Similarly, the provision of storage or compute or network infrastructure may be closely tied to research activities, or also applied to education and administration.

While acknowledging that research depends on the entire support-base across the spectrum, it blurs the measurement, and deeper analysis, of the research-support sector. Not just the numbers but more broadly their career paths, job descriptions and skills-needs are tied specifically to their work with researchers on a regular basis. As such, the survey focussed on those people that would work 'closely' ('at-elbow') with researchers, frequently or infrequently.

Definitions

The IT sector broadly, and the eResearch-support part of it, is rich with a variety of 'formal' role definitions, with an equally wide variety of standards that are applied. When trying to measure the scale of the workforce, we needed to be clear who we are counting and how we are classifying. It is too easy to use a particular set of semantic terms, and then have one respondent apply those terms differently to another.

This is particularly the case around data management, where the distinctions between manager, custodian, curator, steward and other titles are argued across the literature, and various models of data governance are expounded. This also extends to data analysts, scientists and engineers, titles that have widely varying definitions and roles within research processes. The survey used language that described activities, rather than job titles.

One functional distinction was included in the survey, around data responsibilities and the relationship between the data 'caretaker' to the data. There are many people known to be taking care of data that is part of their project, but who do not deal with any other data. At the same time, there is a community of people who take care of data that they are not connected with, and may span multiple projects or even disciplines. The skillsets overlap, but the roles are different. The survey questions separated these functions. An extension of this question is whether such support is only for the duration of a particular project, or extends e.g. into perpetuity, providing archival support. It was decided that the meaning of timeframes varies by disciplines and boundaries may not be clear enough, so questions around duration of data support were left out of this survey.

There is a similar definitional situation emerging around software engineering, where there is some support to formalise and acknowledge the role of "research software engineer", although not many students writing code for their project would perhaps identify as such. Conversely, network engineers have had a variety of industry certifications for many years, where there is formal assessment and recognition. The survey was focussed on the development, maintenance and integration of software as part of a research process. It is worth noting there are edge cases where some software used in research is used for other purposes, or is developed outside of projects as research projects in their own right, but is assumed to be a small fraction of the effort.

Some workforce analyses use role-descriptors, while others use skill-requirements to deliver a particular function. Both of those approaches appear sensible, but reflect different aspects – 'roles' outline what a person is expected to do, while 'skills' target the kind of person desired for a 'role'. Neither appear to capture the actual experience of many in research support, where people extend

or teach themselves into a position or function, or where they 'volunteer' their effort (e.g. PhD students) for projects and were never formally recruited for that role.

The survey focussed on what people are actively working on, regardless of title or role-description. Improving clarity of definitions is a workforce management issue for the hosting organisation, not for the survey, though it may inform such discussions.

The survey also recognised that not all research support is about building or running or managing things, but also the human support aspect, in areas such as advice, guidance, analysis, education and training. An extra question was included in each section to measure that community.

Coverage and measurements

It is widely acknowledged that many people working in a variety of research support roles are not formally funded or employed to do so. A significant amount of research support comes from researchers themselves, colleagues helping each other, and especially PhD students. The survey asked respondents to estimate the count of 'unfunded' staff/students working in each of the topic areas, in addition to the funded positions.

A similar challenge arises with staff who are employed or contribute some fraction of their time towards supporting research. The survey questions encouraged both counting 'people' as well as estimating the equivalent-full-time (EFT) count. This avoids the effort of trying to count individuals at varying fractions, and we still gain a meaningful insight into overall staffing levels and arrangements.

Distribution and collection

Distributing the survey and collecting the responses, within reasonable timeframes, required an active approach. While much of the research sector of interest is university-centred, there is rarely a single person at any university who knows the full extent of research support within their institution. It is possible to identify common touch points, e.g. DVCs-R, IT-Directors, Librarians, and also eResearch Directors where they exist. However, their overview of the entire institution varies immensely. It was decided to engage all the contact points at every university, either directly (DVC-R, eResearch directors) or through national bodies such as CAUL (Librarians) and CAUDIT (IT Directors). In each case they were asked to both report what they knew, and also to encourage participation in the survey widely within their institution.

Engagement with research 'centres', broadly defined, was within and across universities that invest in their own support structures and include both formal employees and informal (e.g. student) contributions. These may include the Australian Research Council (ARC) Centres of Excellence, Cooperative Research Centres (CRC), National Collaborative Research Infrastructure Strategy (NCRIS) capabilities, and other significant research groups. This could be done through discipline groups, such as the NCRIS capabilities, through state eResearch bodies who support them, and associations such as the Australasian Association for Digital Humanities (aa-DH), the Australasian eResearch Organisations (AeRO), and the CRC Association. At a higher level, groups such as the Learned Academies as well as Science and Technology Australia (STA) may provide additional pathways to some significant research groups. Given the timeframe for the survey was short only a few centres were contacted directly, to at least gather some initial data.

Outside of universities there are a range of state and national research support organisations, including e.g. the state eResearch partnerships, national supercomputing centres, network and identity providers, amongst others. These were approached directly about their own workforce, and also for leads into significant institutional/discipline centres.

Beyond the typical higher-education research organisations and their support frameworks are a wide range of other research support providers, especially as data providers, for example government agencies, and the galleries, libraries, archives and museums sector (often written as GLAM). Engaging with these organisations could be through a variety of channels, including direct approaches and also aggregates such as the various Federal and also State Agency IT Leadership groups. Again, time did not permit comprehensive communication with them, but is recommended for a future follow-up survey.

Finally, a key challenge is not knowing what we don't know. There are certainly people working in research support roles that should be part of the survey, but through unknown arrangements, within unknown/unasked groups, with unknown titles or functions. While a top-down survey approach provides aggregated data, and relatively easily, it will not find potentially important groups and emerging roles. To help alleviate this a similar survey should also support a bottom-up approach, where individuals can self-identify and add their own information to the survey. An added benefit is that they may provide 'ground-truth' information for the top-down reports where they overlap. Engaging with that broader community will take much more time and effort, which was not available for this survey, but is strongly encouraged for a follow-up survey. The survey is included in the Appendix.

Result summary

Responses received

As noted in the survey communication, all responses were de-identified and aggregated for the purposes of this report, as far as possible. There were some perceptions of the data being sensitive, so communicating the protection of the data was important.

A total of **42 responses** were received, from:

- 18 universities (providing 29 responses)
    - Of the 18 universities,
        - 5 were from the Group of 8 members,
        - 3 from the Regional Universities Network (RUN) members, and
        - 10 were other major-city based universities
    - Of the 29 responses,
        - 8 were estimates of the entire institution
        - 8 were discipline-centres, faculties, etc, and
        - 13 were functional areas (Libraries, IT, eResearch, etc…)
- 4 state and national "eResearch bodies".
- 9 discipline groups, centres, NCRIS entities, archives, etc. Of those, four came from the medical sector.

Analysis overview

A simple addition of all the reported staffing levels would provide meaningful baseline totals but be difficult to extrapolate to some meaningful national figures. By considering each of the organisational categories separately it is possible to analyse the coverage in each category and derive a somewhat more meaningful extrapolation. The detailed figures in the tables below include both the raw reported figures, as well as the scaling as outlined here.

The **state and national "eResearch bodies"** can be enumerated, so their proportion of responses were calculated, and their figures scaled. There are approximately 8-10 such bodies nationally, of varying size and functions. The four that submitted data represent approximately half of the sector and span the diversity of functions, so a multiplier of 2 seems a reasonable extrapolation.

The **discipline groups, centres**, etc. can only be counted as reported, with no scaling easily derived without a national catalogue of such bodies. In the survey they contributed around 5-15% of the reported totals in each category (expanded further below), and is certainly a gross under-count of their contribution across the sector. It reflects the lack of reach of the survey into that sector, and that many such centres may span multiple universities, and more even sit entirely outside of universities. This is an opportunity for future enhancement of the survey.

The analysis for **universities** is more complex. There are three response types in their case:

- Whole of institution
- Part(s) of institutions
- A combination

Where responses were described as 'whole of institution', and no other data was provided from elsewhere within the institution, the response can be taken at face value. It is worth noting that these responses were generally at the high-end of the numeric scale, and often open-ended. Comments from respondents indicated they were conservative estimates, by well-informed staff, but no university had a single aggregated view of their institution.

In 2 cases the survey received an institution-wide estimate as well as some centres and functional areas within the institution. These were checked for compatibility and/or overlap, and with neither apparently being an issue were combined for their totals.

Most responses were from individual areas, faculties or departments (including IT and Library), where the measurement is likely to be much more reliable. Given only a part of the institution ideally the figures should be scaled to the whole institution. The difficulty though is that any particular research centre may not be a fair representation of the institution, and it is more likely to introduce a bias as active centres may be more willing to respond. It is also difficult to compare figure from e.g. Libraries with those from e.g. a Science faculty. In only one case did an institution provide more than two responses. As such the provided figures are only used directly, to provide a baseline value, on the understanding that the figures are going to be actually (much) higher.

Finally, to **generate national totals** across all response categories given that only a known subset of universities replied, we need to extrapolate the figures accordingly.

For universities it seems reasonable to assume that investment in research support will be somewhat correlated with institutional research investment/income. This assumption could be better analysed with more data and time, but appears a defensible starting point for scaling.

The Department of Education annually publishes[1] a range of research intensity metrics, which in turn are aggregated to determine annual Research Block Grant (RBG) allocations. The universities that responded to this survey here in aggregate covered 61% of the national RBG allocations. If the university responses to this survey had all been for the entire institution a multiplier of around 1.6 would be a reasonable, and probably very conservative extrapolation. As only 8 of the 18 indicated

[1] https://www.education.gov.au/news/2019-research-block-grant-allocations-released

they had reported on the whole institution, and the rest are a combination of just a few centres, faculties and departments within the university, another rough doubling could be reasonably, and still conservatively applied, i.e. a multiplier of around 3.2.

The survey used a range selection to make it easier for respondents to report. For the calculations the mid-point of the range was used. Where '50+' was reported a value of 55 is somewhat arbitrarily used, but since it is likely to be equivalent to an 'educated guess' it is better to remain conservative.

It is known and accepted that the reported figures are conservative and, in many cases, estimates, especially given the obvious under-reporting. This is due in part to the lack of data across the sector of where people are doing this work. A few submissions noted overlap in positions/roles, and their attempts to adjust the figures, but these are unlikely to dramatically change the resultant totals.

Analysis outcomes

The following tables summarise the survey results, showing both the raw data, straight out of the survey, and the scaled figures for the estimated national totals, as proposed above. The scaled figures are rounded to the nearest ten (or 5) for ease of reading. Each table represents one role description from the survey, with columns for the count of people, the equivalent-full-time (EFT) figure, and an estimate of the number (not EFT) of unfunded people, e.g. researchers/students doing the work despite it not being their main (paid) job. The tables break out the three major categories of organisations identified in the survey, i.e. universities, eResearch bodies at state and national levels, and discipline centres such as research institutes, NCRIS capabilities, community archives, etc.

Research Data Support

It is worth noting that research-data roles have gained high visibility in institutions over the last decade through the efforts of ANDS, and now increasingly 'big data' and 'machine learning' initiatives for research at many institutions.

How many people support the…

… Collection/generation of digital data and/or metadata

	People (raw)	**EFT (raw)**	**Unfunded people (raw)**	***People (scaled)***	***EFT (scaled)***	***Unfunded people (scaled)***
eR bodies	10.5	10.5	0	*20*	*20*	*0*
Centres	46.5	28.5	17.5	*50*	*30*	*20*
Universities	284	203.5	148.5	*450-900*	*325-650*	*240-480*
Totals	**341**	**242.5**	**166**	***520-970***	***375-700***	***260-500***

… Analysis of research data, e.g. through modelling, visualisation, machine learning, or other tools

	People (raw)	**EFT (raw)**	**Unfunded (raw)**	***People (scaled)***	***EFT (scaled)***	***Unfunded (scaled)***
eR bodies	34.5	34.5	0	*70*	*70*	*0*
Centres	38.5	32	1	*40*	*30*	*0*
Universities	336	208.5	140.5	*540-1080*	*330-660*	*220-440*
Totals	**409**	**275**	**141.5**	***650-1190***	***430-760***	***240-460***

… Management of digital data for a project they are closely involved with (e.g. Data Managers, Custodians)

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	4.5	4.5	0	*10*	*10*	*0*
Centres	18	15.5	1	*20*	*20*	*0*
Universities	266.5	179	152	*430-860*	*290-580*	*240-480*
Totals	**289**	**199**	**153**	***460-890***	***320-610***	***240-480***

… Stewardship of digital data on behalf of others (e.g. supporting repositories or data governance)

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	9.5	9.5	0	*20*	*20*	*0*
Centres	24	19.5	4.5	*20*	*20*	*5*
Universities	278.5	158.5	11.5	*450-900*	*250-500*	*20-40*
Totals	**312**	**187.5**	**16**	***490-940***	***290-540***	***25-45***

… Provision of advice, education/training in managing digital data and/or metadata

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	45	15	0	*90*	*30*	*0*
Centres	24	16	3.5	*20*	*20*	*5*
Universities	225	119	27.5	*360-720*	*190-380*	*40-80*
Totals	**294**	**150**	**31**	***470-830***	***240-430***	***45-85***

The figures show an average 60-70% EFT load compared to the people count in most organisations, suggesting that a large fraction of the people may be spread across multiple functions, or a level of part-time effort. Summing the people counts across all functions may lead to double-counting, but in principle we can add the EFT figures, to get a sense of the total data-support workforce.

Interestingly, the unfunded people counts, likely to be researchers and/or HDR students, are of a similar order to the funded people counts for the first three functions (i.e. excluding stewardship and training), despite being very likely a conservative estimate.

If the higher end of the extrapolation is used to minimise the conservative reports, the data suggests that:

- around 2000 people (1500 EFT) are employed in the collection and analysis of research data,
- around 2000 people (1100 EFT) are employed in its management or stewardship, and
- around 800 people (400 EFT) are employed in training/advising researchers on research data

It is useful to compare this to the scale of the community they support. The national ‘research and development workforce’ (staff, excluding students) is estimated by the ABS at around 40,000

(person-year-equivalent) in 2015-2016, across higher education[2] plus government and non-profits[3], who tend to collaborate with universities. There is also data relating to businesses[4], who may in many cases also collaborate with university researchers, adding potentially up to another 33,000. Without further information though it is difficult to calculate their contribution with any precision. Analysing the data from the Department of Education[5] suggest a Higher-Degree-by-Research (HDR) student cohort of around 45,000 in 2017, in good agreement with the ABS report.

Taking all of these into account and assuming a total research workforce of perhaps around 100,000, the results indicate that it is supported at roughly:

- 1 EFT per 60-70 researchers for collection and analysis of research data,
- 1 EFT per 90 researchers for management or stewardship of research data, and
- 1 EFT per 250 researchers providing training/advice on research data.

There are no support-quality measures in this survey to indicate how well the research community is served by this support. It does directly suggest there is a sizable, and probably growing, community of support staff that needs to be developed, retained, sufficiently skilled, perhaps accredited, and valued**.**

Research Software Support

In comparison to research-data, the support of research software, its development, engineering, maintenance and support, still has low formal visibility in institutions, and even less clarity around its role definitions. This is a global phenomenon and well-known challenge. Only over the last few years has the 'Research Software Engineer' (RSE) description become more visible, and it still struggles to capture the diversity of relationships between researchers and software.

How many people …

… Develop/engineer/maintain/support software used in research analysis

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	45.5	45.5	0	*90*	*90*	*0*
Centres	48	48	8	*50*	*50*	*10*
Universities	268.5	239.5	124.5	*430-860*	*380-760*	*200-400*
Totals	**362**	**333**	**132.5**	***570-1000***	***520-900***	***210-410***

[2] https://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/AE02B963FB1D51B2CA2571B60075B1C0?opendocument - this combines post-graduate, researchers and 'other supporting staff'.

[3] https://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/0AE638AFEA290E1BCA256964007CF648?opendocument - this combines researchers with technical staff and 'other staff'

[4] https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8104.0Main+Features12015-16?OpenDocument - which notes around 33,000 'researchers' but no reference to other roles.

[5] https://docs.education.gov.au/node/47831

… Provide advice, education/training to researchers/research teams on relevant tools and applications

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	55	42.5	0	*110*	*80*	*0*
Centres	21.5	13.5	4.5	*20*	*10*	*5*
Universities	187.5	121.5	21.5	*300-600*	*200-400*	*30-60*
Totals	**264**	**177.5**	**26**	***430-730***	***290-490***	***35-65***

Interestingly, the EFT fraction reported for software development is around 90%, which could indicate these roles are more clearly bounded, or perhaps more likely, has failed to capture a large pool of part-time/casual people working in this space. Training though is once again significantly on a more fractional basis.

Taking again the higher end of the scaled ranges, the data indicates at least 1000 people (over 900 EFT) are involved nationally in various 'software-engineering' or similar development related roles, or about 1 EFT per 100 researchers. Based on many years of observations of research groups across many institutions it seems likely that this figure is a significant under-estimate, of both funded and unfunded roles.

Further, it suggests 500 EFT are providing training and advice, or 1 EFT per 200 researchers, which is unlikely to capture the many students and early-career researchers informally teaching their supervisors and/or teams.

Research IT Infrastructure Support

As outlined in the introduction, there is a wide spectrum of IT infrastructure and its provision across universities that supports research, and also potentially other activities. The question attempted to focus the measurement on those people supporting infrastructure that supported research directly or closely.

How many people …

… Build/operate local IT infrastructure (e.g. computing platforms, storage services, specialised networks) used primarily for research

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	41.5	41.5	0	*80*	*80*	*0*
Centres	21	16.5	0	*20*	*20*	*0*
Universities	160	137.5	38.5	*260-520*	*220-440*	*60-120*
Totals	**222.5**	**195.5**	**38.5**	***360-620***	***320-540***	***60-120***

… Support access to external IT infrastructure (e.g. computing platforms, storage services)?

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	18.5	18.5	0	*40*	*40*	*0*
Centres	15	10.5	0	*20*	*10*	*5*
Universities	107.5	86.5	20.5	*170-340*	*140-280*	*30-60*
Totals	**141**	**115.5**	**20.5**	***230-400***	***190-330***	***35-65***

… Provide advice, education/training to researchers/research teams on research IT infrastructure

	People (raw)	EFT (raw)	Unfunded (raw)	*People (scaled)*	*EFT (scaled)*	*Unfunded (scaled)*
eR bodies	41	38.5	0	*80*	*80*	*0*
Centres	10.5	6	0	*10*	*5*	*0*
Universities	140	95.5	11.5	*220-440*	*160-320*	*20-40*
Totals	**191.5**	**140**	**11.5**	***310-530***	***245-405***	***20-40***

The data shows a strong contribution at universities and eResearch bodies, as could be expected, with a reasonably high proportion of EFT to people-count, at around 80-85% of 1000 staff and another 500 involved in training. This perhaps reflects that most of the 'significant' infrastructure is provided at institutional scale, with allocated staff. However, the small proportion of unfunded staff may suggest the survey has not reached the many staff and students supporting project-scale systems for data storage and compute platforms, which are known to exist. Of note also the people/EFT counts are distinctly smaller than those for the data and software support questions. Again, this may just reflect the scalability achieved with institutional infrastructure and perhaps also neglects project infrastructure.

Training support is quite comparable across all categories though. There is a stronger component in these figures from the eResearch bodies, many of whom host e.g. high-performance computing platforms and storage systems and provide relevant training to maximise usage.

Respondent Commentary

Survey respondents were asked to add free-text comments in relation to individual questions and to the survey overall.

The most common comment by far came from senior levels in universities, and could be summarised as "frankly, outside of X, we do not know". No university indicated they had any detailed measures of what happened outside the central areas, i.e. beyond IT, Library and eResearch programs, and this was common from the largest to the smallest institution. They all indicated their estimates were known to be conservative. One senior executive asked if they could be provided with the data submitted by their own institution. Several acknowledged the importance of a better understanding.

Quite a few respondents indicated that "it's complicated". Some said that they would establish teams for particular projects that evolved as required, and staff were then redirected between projects, but no comments were made about any risks that may arise with that approach. Many also tried to explain in detail how the various fractional roles combined, and how perhaps internal and external resources were called in as needed, both formally and informally, e.g. training in statistics software provided by researchers in another discipline.

A couple of responses noted the questions were perhaps not clear enough in what they were measuring, if these were counting 'specialist' staff, dedicated to a purpose or more broadly. These responses said that some of these functions were provided through business-as-usual functions established in e.g. the Library, IT Department or Research Office. These may be useful input for future survey design.

Some of the individual questions had extra comments to identify staff involved in research support in other, but variously related functions. The areas noted included legal, ethics and privacy support,

where data had to be shared during a project; literature searches for projects undertaking e.g. surveys - a common process for clinical trials; policy and strategy development including eResearch-related committees; and assistance with data searches across library databases. One flagged that collection-of-data did not distinguish those collecting data from sensors, and those building/deploying the sensors that actually collected the data.

One University submitted a null response to all questions, and just added in effect “we have no data at all, but are establishing a strategy, recognise the importance, and are keen to be involved in discussions”. This submission was not counted in the totals above.

A future survey may wish to allow for fractional (i.e. 0-1) counts of EFT. Three centre-based responses basically said: “it’s all just me! And sometimes I get help”.

Closing

This survey has attempted to provide the first broad-scale overview of the national research IT support workforce, and is perhaps the first of its type anywhere. While the figures appear mostly very conservative, they do indicate the scale and breadth of the workforce that is already out there, with well over 6000 staff in support roles and 2000 providing training/advice, and demand is clearly growing. Given the number of survey responses in a short time and with fairly minimal promotion, and the not-insignificant effort put into many of those responses, there is clearly great interest in the sector to understand more about the workforce. It is hoped that this type of survey can be repeated, and enhanced, in the future.

Appendix - The Survey and its implementation:

The survey was developed using surveymonkey.com as the platform.

- The questions were deliberately made very simple, avoiding semantic terms that required interpretation.
- The overall question count was limited and kept to a single page, to make it less intimidating to respondents.
- Questions used a matrix-dropdown approach, effectively one question section for each given area (data, software, infrastructure), broken into a small number of short questions relating to people involved in each of the relevant functions.
- The dropdown responses provided an easy numeric input, using ranges ('1', '2-3','4-5', …, '30-40', '40-50', '50+'), under headings of 'people', 'EFT' and 'unfunded'. This was intended to remove the urge to count to high precision, and was unlikely to have a material effect on the final tallies.
- Each question included some explanatory text at the front, and an open text field at the end for comments to be provided. There was also a final 'any other comments' opportunity.

The survey was established with a readable URL (https://www.surveymonkey.com/r/ARDC-workforce) to make it easy to share via emails, social-media, etc. An email summary of the survey intentions was developed to invite participation for direct contacts, and to be shared, with any modifications, through associations and other groups. The email was reflected in the survey introduction. From the initial announcement a time-limit of 3 weeks was set for responses, although a few arrived late, usually accompanied by a phone call asking for permission.

[ARDC Logo]

How big is the Research Support Workforce?

This survey aims to obtain an accurate indication of the size of the national workforce involved in supporting digital data, research software and associated infrastructure. It is being carried out by the Australian Research Data Commons (ARDC) to support planning on future workforce needs in this sector, and in preparation for the eResearch Skilled Workforce Summit in Sydney 29-30 July 2019.

We are seeking responses from across the sector, including research institutions and supporting eInfrastructure providers, such as universities, research centres, Publicly Funded Research Agencies (PFRAs), regional/state/national support organisations, etc.

You may:

- respond for an institution, organisation, group, etc. that you lead, oversee or work with, or
- respond collectively for multiple separate organisations/groups that you know well, aggregating the people counts as appropriate, as long as you can identify them below.

You can also do the survey multiple times for different groups if you wish.

The analysis will work to remove duplication if multiple responses happen to be received from a single group or organisation, or where people may do multiple roles.

Please provide your best estimates for each of the questions below, and err on the side of including, rather than excluding people in your counts. The questions are designed to be easy for quick estimated responses.

There are spaces to provide additional information or clarification.

Note: all responses will be anonymised and aggregated for the final report. ARDC will release the final report publicly in July 2019.

If you have any questions, please contact the survey team via email or contact ARDC via the links on the survey announcement page.

1. A bit about you and who you are responding for

Name

Where are you from

Organisation(s)/Group(s), etc. you are responding for

Your email address

2. Are you willing to be contacted if we need any clarification about your responses?

Yes/No

The following questions are about the support of research data.

These include a very wide variety of roles with many different definitions and titles, but their core function is centred on research data collection, analysis and management. This may be for projects they are working in (as staff or collaborators), or as stewards where they are entrusted with data by others. They may include staff with appropriate qualifications formally employed to undertake these roles, through to researchers and students contributing their time to progress a project.

We want to count people who primarily work with researchers, and may include employees, contractors, interns, researchers and students.

In each case, the main interest is the 'people count'. Try to also provide equivalent-full-time (EFT) totals to allow for better comparisons across the sector. The 'unfunded count' aims to identify people who are doing this work but it's not their main responsibility (like researchers and students).

3. How many people support the...

- Collection/generation of digital data and/or metadata? [People/EFT/Unfunded]
- Analysis of research data, e.g. through modelling, visualisation, machine learning, or other tools? [People/EFT/Unfunded]
- Management of digital data for a project they are closely involved with? (e.g. Data Managers, Custodians) [People/EFT/Unfunded]
- Stewardship of digital data on behalf of others (e.g. supporting repositories or data governance) [People/EFT/Unfunded]
- Provision of advice, education/training in managing digital data and/or metadata [People/EFT/Unfunded]

Are there people in other research data roles not identified above? If yes, what are they doing, and how many are there (people, EFT, unfunded)

The following questions are about development, operation and training around research-specific software.

This can include staff with computing skills who formally support researchers in their development and/or use of software tools; those who develop/maintain software for researchers; researchers/students who do some programming to progress a research project (either their own or for others); and researchers/students who undertake systems administration while it is not a formal part of their job description.

We want to count people who primarily work with researchers, and may include employees, contractors, interns, researchers and students.

In each case, the main interest is the 'people count'. Try to also provide equivalent-full-time (EFT) totals to allow for better comparisons across the sector. The 'unfunded count' aims to identify people who are doing this work but it's not their main responsibility (like researchers and students).

4. How many people...

- Develop/engineer/maintain/support software used in research analysis [People/EFT/Unfunded]
- Provide advice, education/training to researchers/research teams on relevant tools and applications? [People/EFT/Unfunded]

Do you have people in other research software roles not identified above? If yes, what are they doing, and how many are there (people, EFT, unfunded)?

The following questions are about development, operation and training around research-specific IT hardware.

This may include staff with formal qualifications who provision/operate research-specific computing, storage or networking hardware, through to researchers/students supporting projects themselves.

We want to count people who primarily work with researchers, and may include employees, contractors, interns, researchers and students. If people provide multiple functions below please distribute their count as seems most appropriate to you.

In each case, the main interest is the 'people count'. Try to also provide equivalent-full-time (EFT) totals to allow for better comparisons across the sector. The 'unfunded count' aims to identify people who are doing this work but it's not their main responsibility (like researchers and students).

5. How many people...

- Build/operate local IT infrastructure (e.g. computing platforms, storage services, specialised networks) used primarily for research? [People/EFT/Unfunded]
- Support access to external IT infrastructure (e.g. computing platforms, storage services)? [People/EFT/Unfunded]
- Provide advice, education/training to researchers/research teams on research IT infrastructure? [People/EFT/Unfunded]

Do you have people in other research IT hardware roles not identified above? If yes, what are they doing, and how many are there? (people, EFT, unfunded)

6. Do you have any additional comments or information?